

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



## **Modelação das Zonas Tarifárias no Seguro Automóvel**

Andreia Alexandra Ribeiro Ginja

**Mestrado em Matemática Aplicada à Economia e Gestão**

Trabalho de Projeto orientado por:  
Maria Isabel Calisto Frade Barão  
Luís Manuel Maranhão Ferreira

2017



## Agradecimentos

Ao meu orientador da empresa Luís Manuel Maranhão Ferreira por toda a disponibilidade e interesse demonstrado.

À minha colega de trabalho Joana Garraio pela grande paciência e disponibilidade para me ajudar sempre que eu precisava.

Aos meus pais por todo o apoio e principalmente porque foram eles que me possibilitaram tirar este mestrado em Matemática Aplicada à Economia e Gestão e assim aprofundar ainda mais os meus conhecimentos em Matemática.

Um agradecimento especial à professora Teresa Alpuim e à professora Maria Isabel Barão.

E a todos os meus amigos, família, colegas de equipa do Porto e ao César pelo grande apoio, convívio e motivação que me deram para eu pudesse concluir este projeto.

Espero que após esta grande etapa, consiga de alguma forma, retribuir e compensar todo o apoio e dedicação que me deram. A todos, um grande beijinho e obrigada!



## Resumo

O presente projeto tem como objetivo criar e modelar as zonas tarifárias de Portugal Continental, no seguro automóvel, usando o modelo linear generalizado (GLM) e encontrar um cliente-alvo para focus do negócio. Como primeiro passo foi definido que apenas iria ser utilizado como base de estudo Portugal Continental devido à falta de informação das apólices cujos clientes são residentes nas ilhas da Madeira e Açores.

Foi utilizada como estatística principal do projeto uma medida muito importante de uma carteira de Seguros, a frequência de sinistralidade, proporção de sinistros por exposição ao risco.

Foram analisados dados referentes aos últimos três anos completos (2013, 2014 e 2015) para ter uma visão mais atualizada e vasta da frequência dos sinistros automóveis para se poderem tirar conclusões mais assertivas. Contudo alguns dos valores foram corrigidos manualmente porque a base usada tinha alguns erros de classificação.

Foi construído um simulador para uma melhor perceção de como a frequência real de sinistralidade se comporta, em diferentes segmentos, de alguns fatores de risco.

Após a modelação das zonas tarifárias foi criado um cliente-alvo, ou seja, através do novo modelo procurou-se descobrir quais seriam os segmentos referentes às características do cliente com menor frequência de sinistralidade.

**Palavras-chave:** Seguro Automóvel, Zonas Tarifárias, Modelo Linear Generalizado, Frequência, Sinistralidade

## **Abstract**

The objective of this project is to model the tariff zones of Portugal, in the motor insurance, using generalized linear model (GLM). As a first step it was decided that only Portugal Continental would be studied due to lack of information about clients residing on the islands of Madeira and Azores.

A very important measure in insurance- the frequency, proportion of the number of claims per the exposure to risk- was use as the main statistic of the project.

Data was analyzed for the last three full years (2013, 2014 and 2015) to have a wider vision of the frequency of claims and to take more assertive conclusions. However some of the values were corrected manually because the database used had some misclassification.

A simulator was built for a better perception of how the real claims' frequency behaves in different segments of some risk factors.

After modeling the tariff zones, target-customer was created, that is, through the new model, we tried to find out which segments of the customer's characteristics were referring with less frequency.

**Keywords:** Motor insurance, tariff zones, Generalized Linear Model, frequency, claims

# Índice

Lista de tabelas: .....	2
Lista de esquemas: .....	2
Lista de equações: .....	3
Lista de figuras: .....	3
Lista de Mapas: .....	3
Lista de gráficos: .....	3
Introdução .....	4
Capítulo 1. A Atividade Seguradora .....	5
1.1 Grupo AXA .....	5
1.2 AXA Portugal, Companhia de Seguros SA.....	5
1.2.2 Estrutura Organizacional .....	6
1.3 Contrato de Seguro.....	7
1.3.1 Conceito .....	7
1.3.2 Ramos de Seguro.....	9
1.3.3 Tarifa .....	13
Capítulo 2. Enquadramento Teórico .....	16
2.1 Frequência: Conceito.....	16
2.2 Modelo Linear Generalizado.....	18
2.2.1 Definição .....	18
2.2.2 Família Exponencial.....	20
2.3 Métodos de Correlação.....	22
2.4 Seleção de Variáveis .....	24
2.4.1 Métodos de seleção .....	25
2.4.2 AICc .....	29
Capítulo 3. Caso de Estudo: Portugal Continental .....	30
3.1 Preparação dos dados .....	30
3.2 Frequência .....	31
3.2.1 Frequência Real.....	31
3.2.2 Frequência Ponderada .....	38
3.2.3 Frequência Conjunta .....	41
3.3 Modelação .....	44
3.4 Modelo Final .....	67
Conclusão/Discussão.....	80
Referências Bibliográficas .....	82
Referências Sitográficas .....	82

## Lista de tabelas:

**Tabela 2.1.** Exemplo Prático

**Tabela 2.2.** Resultados do exemplo prático

**Tabela 2.3.** Distribuições da família exponencial

**Tabela 2.4.** Modelo Lineares Generalizados

**Tabela 3.3.** Descrição das variáveis

**Tabela 3.2.** Legenda da Frequência

**Tabela 3.3.** Correspondência da freguesia (Lisboa)

**Tabela 3.4.** Correspondência da freguesia (Porto)

**Tabela 3.5.** Correspondência da freguesia (Vila Nova de Gaia)

**Tabela 3.6.** Percentagem de concelho de Portugal Continental por zona tarifária

**Tabela 3.7.** Escolha dos métodos

**Tabela 3.8.** Estimativa dos coeficientes da variável bonus

**Tabela 3.9.** Estimativa dos coeficientes da variável zonarede

**Tabela 3.10.** Estimativa dos coeficientes da variável codcomb

**Tabela 3.11.** Estimativa dos coeficientes da variável idadecond

**Tabela 3.12.** Estimativa dos coeficientes da variável idadecart

**Tabela 3.13.** Estimativa dos coeficientes da variável contrato

**Tabela 3.14.** Estimativa dos coeficientes da variável fpagamento

**Tabela 3.15.** Estimativa dos coeficientes da variável cliente

**Tabela 3.16.** Estimativa dos coeficientes da variável pack

**Tabela 3.17.** Estimativa dos coeficientes da variável marca (1)

**Tabela 3.18.** Estimativa dos coeficientes da variável marca (2)

**Tabela 3.19.** Estimativa dos coeficientes da variável marca (3)

**Tabela 3.20.** Estimativa dos coeficientes da variável concelho2 (1)

**Tabela 3.21.** Estimativa dos coeficientes da variável concelho2 (2)

**Tabela 3.22.** Estimativa dos coeficientes da variável concelho2 (3)

**Tabela 3.23.** Estimativa dos coeficientes da variável concelho2 (4)

**Tabela 3.24.** Estimativa dos coeficientes da variável concelho2 (5)

**Tabela 3.25.** Estimativa dos coeficientes da variável concelho2 (6)

**Tabela 3.26.** Estimativa dos coeficientes da variável concelho2 (7)

**Tabela 3.27.** Estimativa dos coeficientes da variável concelho2 (8)

**Tabela 3.28.** Estimativa dos coeficientes da variável concelho2 (9)

**Tabela 3.29.** Estimativa dos coeficientes da variável concelho2 (10)

**Tabela 3.30.** Estimativa dos coeficientes da variável concelho2 (11)

**Tabela 3.31.** Estimativa dos coeficientes da variável concelho2 (12)

**Tabela 3.32.** Estimativa dos coeficientes da variável concelho2 (13)

**Tabela 3.33.** Estimativa dos coeficientes da variável concelho2 (14)

**Tabela 3.34.** Estimativas dos coeficientes da variável CVE\_TIPO\_GARAGEM

**Tabela 3.35.** Estimativa dos coeficientes da variável idadeveic

**Tabela 3.36.** Estimativa dos coeficientes da variável NUM\_CC\_categ

**Tabela 3.37.** Estimativa dos coeficientes da variável NUM\_TARA\_categ

## Lista de esquemas:

**Esquema 1.1** Características do contrato de seguro

**Esquema 2.2.** Método Forward

**Esquema 2.3.** Método Backward

**Esquema 2.4.** Método Stepwise

**Esquema 3.5.** Frequência Ponderada

**Esquema 3.2.** Frequência Conjunta



## Lista de equações:

**Equação 2.1** Frequência de Sinistros

**Equação 2.2.** Cálculo da exposição

**Equação 2.3** Coeficiente  $R^2$

**Equação 2.4.** Cálculo do AICc

**Equação 3.1.** Modelo Final

## Lista de figuras:

**Figura 1.1** Garantias das Coberturas do Seguro de Danos Próprios

**Figura 1.2.** Seguros Especiais

**Figura 2.3.** Correlações entre as variáveis sem qualquer limite

**Figura 2.4.** Correlações entre variáveis acima de 0.06

**Figura 2.5.** Correlações entre variáveis acima de 0.2

**Figura 3.6.** Variáveis do Simulador

**Figura 3.7.** Correlações entre as variáveis

**Figura 3.8.** Correlações entre as variáveis com um limite de correlação igual ou superior a 0.2

**Figura 3.9.** Correlações entre as variáveis com um limite de correlação igual ou superior a 0.7

## Lista de Mapas:

**Mapa 3.1.** Portugal Continental (frequência real)

**Mapa 3.2.** Concelho de Lisboa repartido por freguesias (frequência real)

**Mapa 3.3.** Concelho do Porto repartido por freguesias (frequência real)

**Mapa 3.4.** Concelho de Vila Nova de Gaia repartido por freguesias (frequência real)

**Mapa 3.5.** Evolução anual da frequência real (Portugal Continental - concelhos)

**Mapa 3.6.** Simulador de Frequência real de sinistralidade

**Mapa 3.7.** Portugal Continental (frequência ponderada)

**Mapa 3.8.** Evolução anual da frequência ponderada (Portugal Continental – concelhos)

**Mapa 3.9.** Portugal Continental (frequência conjunta)

**Mapa 3.10.** Evolução anual da frequência conjunta (Portugal Continental – concelhos)

## Lista de gráficos:

**Gráfico 3.1.** Aplicação do modelo na variável bonus

**Gráfico 3.2** Aplicação do modelo na variável zonarede

**Gráfico 3.3** Aplicação do modelo na variável codcomb

**Gráfico 3.4** Aplicação do modelo na variável pack

**Gráfico 3.5** Aplicação do modelo na variável cliente

**Gráfico 3.6** Aplicação do modelo na variável fpagamento

**Gráfico 3.7** Aplicação do modelo na variável idadecond

**Gráfico 3.8** Aplicação do modelo na variável idadecart

**Gráfico 3.9** Aplicação do modelo na variável contrato

**Gráfico 3.10** Aplicação do modelo na variável idadeveic

**Gráfico 3.11** Aplicação do modelo na variável CVE\_TIPO\_GARAGEM

**Gráfico 3.12** Aplicação do modelo na variável NUM\_CC\_CATEG

**Gráfico 3.13** Aplicação do modelo na variável NUM\_TARA\_CATEG

## Introdução

O seguro automóvel é um seguro obrigatório no que respeita à cobertura de Responsabilidade Civil, e tem como finalidade proteger eventuais vítimas de acidentes de viação.

Este seguro responde ao que é legalmente exigido relativamente à obrigação de segurar a responsabilidade civil decorrente da circulação de veículos a motor. Contudo ao longo dos anos, tem sido cada vez mais evidente a concorrência na atividade seguradora, pelo que a área de *pricing* é cada vez mais importante. Na modelação de zonas tarifárias há que ter em conta diversos fatores, como a competitividade com o mercado, futura rentabilidade da empresa, a aceitação do produto por parte do cliente, as características do produto (variáveis tarifárias), a proteção que o produto oferece para o cliente e a relação de preço-qualidade.

Para a criação deste produto final primeiro há que procurar quais os fatores de risco, ou seja, quais são as características do meu cliente que são fundamentais para aplicar no prémio. E é neste último ponto que decorreu o meu projeto. No primeiro passo construí zonas tarifárias com base na frequência de sinistros e posteriormente procurei quais eram as características do meu cliente (modelação) em que existia variação da frequência de sinistros nos concelhos de Portugal Continental.

Neste projeto foram utilizados diversos *softwares*:

- SAS Base 9.1.3 para a construção da base de dados;
- Macro Visual Basics do Excel para a construção das zonas tarifárias;
- Emblem para a construção do modelo.

# **Capítulo 1. A Atividade Seguradora**

No capítulo 1 pretende-se dar a conhecer o mundo segurador, primeiro começando por apresentar a companhia de seguros AXA, no qual foi realizado o estágio curricular e de onde foi recolhida toda a informação presente neste projeto, e posteriormente irão ser apresentada os conceitos da atividade seguradora.

## **1.1 Grupo AXA**

A AXA é uma Companhia de Seguros francesa com atuação sediada em Paris, cuja marca foi oficializada em 1985 graças a Claude Bébéar, e desde esse momento posiciona-se como um dos líderes mundiais da atividade seguradora.

A sua história remota a 1817 quando a AXA foi criada através da fusão de várias companhias que cobriam riscos agrícolas na Normandia. A criação do futuro Grupo AXA é estabelecida sobre a união de 3 ramos de seguro (incêndio, vida e acidentes) que até ao momento eram independentes dando origem ao Grupo Ancienne Mutuelle.

Com cerca de 102 milhões de Clientes em 56 países, a empresa AXA centra-se em três linhas negócio: Não Vida, Vida e Gestão de ativos, contando com 161000 colaboradores e distribuidores.

Desempenhando um papel de destaque nos serviços de proteção financeira, de pessoas e de bens, a AXA tem como principal responsabilidade construir e desenvolver uma sociedade mais forte e mais segura, regendo por servir os seus Clientes tendo como fator guia a Confiança estabelecida entre os dois.

## **1.2 AXA Portugal, Companhia de Seguros SA**

A AXA é uma empresa seguradora com mais de 30 anos de história contudo só no final do século XX surge em Portugal após variadas fusões ao longo dos anos.

### **1.2.1 História**

Para poder falar da história da atividade seguradora em Portugal é preciso primeiro abordar a própria história de Portugal. Portugal é um dos países mais antigos da Europa sendo um dos principais pioneiros da descoberta marítima, e foi em 1293 que D. Dinis rei de Portugal nessa época criou a primeira forma de seguro, inicialmente esta forma de seguro foi dedicada exclusivamente aos riscos marítimos, e por isso mesmo D. Dinis ordenou que fosse celebrado um acordo entre os mercadores, tendo como objetivo o pagamento de certas quantias (“prémio”) sobre as embarcações, para suportar eventuais perdas de navios e mercadorias.

D. Fernando deu continuidade ao trabalho criado pelo seu antecessor tendo sido criada em 1380 a Companhia das Naus que na altura funcionava como uma companhia de seguros. A criação da Companhia das Naus tinha como objetivo proporcionar aos proprietários dos navios uma certa segurança em caso de sinistros, como tal todas as embarcações tinham que estar registadas, pagando uma percentagem dos lucros de cada viagem para a caixa comum que servia para cobrir os prejuízos em caso de afundamento ou avaria. Este foi um passo muito importante na área seguradora em Portugal e em 1383 é publicado a primeira Lei Nacional sobre Seguros.

A área seguradora foi sempre evoluindo ao longo dos séculos, e em 1991 surge a união de quatro companhias a operarem em Portugal, duas delas portuguesas, a Aliança Seguradora, SA e a Companhia de Seguros Garantia, SA (vida e não vida) e outras duas de raiz francesa mas estabelecidas em Portugal como UAP Portugal, Companhia de Seguros, SA que decidem juntar as suas atividades e estabelecem um protocolo para a criação do Grupo Segurador Aliança UAP. Finalmente em 1997, a Aliança UAP muda de nome e de imagem depois da fusão entre o Grupo AXA e a Aliança UAP passando a chamar-se AXA Portugal até ao presente ano de 2016 em que o Grupo AXA decide desvincular a empresa AXA Portugal do grupo e vender a respetiva empresa ao grupo segurador belga Ageas, denominando-se a partir de esse momento Ageas Seguros.

### **1.2.2 Estrutura Organizacional**

Até Março de 2016 inclusive, a estrutura organizacional da AXA Portugal assentava em várias unidades departamentais tal como acontece com todas as empresas do Grupo AXA. Esta divisão visava promover a eficácia e a eficiência de uma organização reconhecida a nível mundial, com uma clara definição de autoridade sendo esta divisão efetuada por áreas de negócio.

#### ***1.2.2.1 DTO – Atuariado Não Vida***

O estágio curricular no qual se baseia este relatório foi realizado no Departamento de Técnica e Oferta mais especificamente no atuariado Não Vida.

A área de técnica e oferta é uma das áreas mais importantes numa empresa seguradora porque é neste departamento que se concebe, se desenvolve e se implementa toda a estratégia para ajustar a oferta e o preço dos produtos que existem na empresa seguradora.

A equipa de atuariado Não Vida está englobada nesta área porque é no Atuariado que se avalia e se administra o risco, é também calculada a probabilidade de ocorrer algum evento futuro, são calculadas as provisões técnicas que representam uma eventual despesa futura para a empresa, isto é, são valores calculados matematicamente com base no histórico de resolução dos sinistros que garantem o cumprimento de uma determinada obrigação por parte da empresa.

O atuário é visto como alguém que de uma forma geral lida com “os números” contudo existe dentro desta área um grande nível de especialização, pois são desempenhados papéis bastante distintos numa companhia de seguros como:

- Pricing: Análise do preço e gestão de portfólio;
- Reserving: Cálculo atuarial de Reservas/Provisões;
- Finance: Análise financeira e contabilística;
- Capital modelling: Estimacão e alocação de capital financeiro de forma a respeitar níveis de solvência.

## **1.3 Contrato de Seguro**

### **1.3.1 Conceito**

É definido como contrato de seguro um acordo celebrado entre uma entidade autorizada a exercer a atividade de seguros (segurador), que se obriga, mediante o recebimento de determinada quantia (prémio), a garantir um determinado risco (situação coberta) e, caso o mesmo se verifique, a pagar à outra parte do contrato (tomador, pessoa segura ou beneficiário) um determinado montante convencionado.

Nos termos do novo regime jurídico do contrato de seguro, os contratos de seguro agrupam-se em dois grandes tipos: seguro de pessoas e seguro de danos (em certa medida, aos ramos vida e não vida, respetivamente). Enquanto o seguro de pessoas compreende a cobertura de riscos relativos à vida, à saúde e à integridade física de uma pessoa ou de um grupo nele identificadas, o seguro de danos respeita a coisas, bens imateriais, créditos e quaisquer outros direitos patrimoniais.

Em cada contrato de seguro existe uma apólice que consiste num documento escrito que confere eficácia e oponibilidade ao contrato celebrado que inclui todo o conteúdo acordado entre o tomador (entidade que fica responsável pelo pagamento do prémio, que também poderá ser a pessoa segura ou o beneficiário) e o segurador (entidade que exerce a atividade seguradora, sendo responsável pela indemnização em caso de sinistro), nomeadamente:

- As condições gerais: cláusulas do contrato que preveem os direitos e as obrigações básicas de ambas as partes, normalmente comuns a todos os contratos que se destinam ao mesmo tipo de risco ou cobertura;
- As condições especiais: cláusulas do contrato que permitem esclarecer ou completar algumas das condições gerais do mesmo contrato, bem como limitar ou ampliar a cobertura em relação às situações e bens previstos nas condições gerais;
- As condições particulares: cláusulas que individualizam o contrato, relativamente à identificação do tomador, do segurado e das pessoas seguras, à duração do contrato, ao montante de prémio a pagar e a outros encargos.

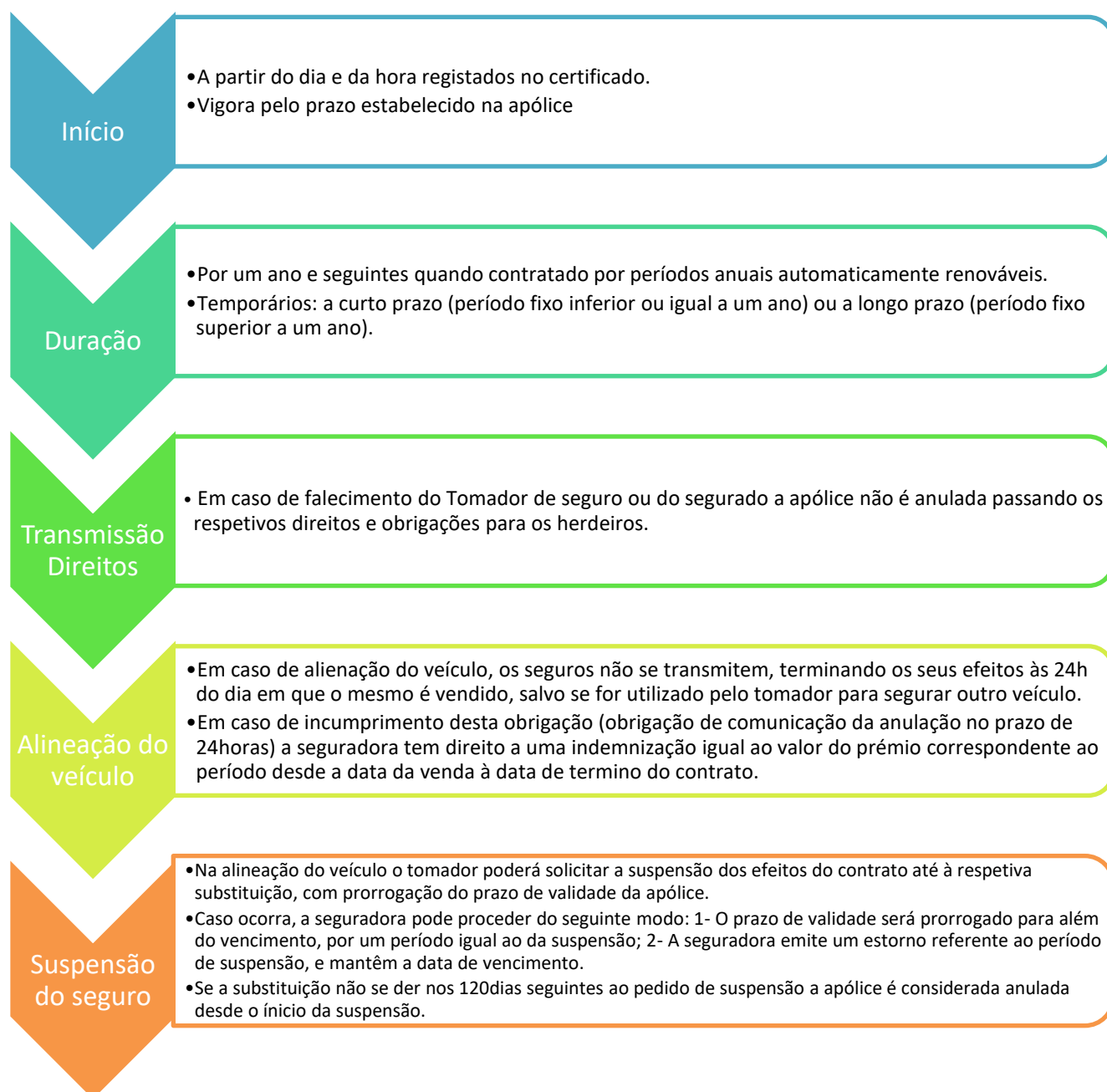
Existem alguns contratos que, pela sua natureza, dimensão ou características, necessitam da intervenção de várias seguradoras. Para tal necessitam de arranjar forma de não colocar a situação financeira em risco, existem dois tipos de contratos, o resseguro e o cosseguro. No primeiro caso, as seguradoras transferem para resseguradores parte do risco que assumiram ao celebrar contratos de seguro com outrem, no segundo caso é um contrato mediante o qual várias seguradoras, de entre os quais um é o líder e sem que haja solidariedade entre eles, assumem conjuntamente um determinado risco, prevendo as mesmas garantias, idêntico período de duração e um prémio global, e constando a quota-parte do risco ou a parte percentual do capital seguro assumido por cada cossegurador.

A palavra risco é bastante usada nos dias de hoje, mas o que significa a palavra “risco” na atividade seguradora? Usualmente o risco tem uma conotação negativa, no entanto o que está subjacente é o nível de incerteza porque o risco representa a possibilidade de ocorrência de um acontecimento incerto (sinistro), alheio à vontade dos intervenientes, contudo previsto na apólice.

Riscos tipicamente determinados por situações causais:

- Não estão relacionados com longevidade, mas podem cobrir situações de morte;
- Seguros que cobrem tanto pessoas como bens;
- Divididos em seguros obrigatórios e facultativos;
- Quando os seguros são “vendidos” diretamente ao segurado trata-se de seguro direto;
- Quando o seguro é partilhado por várias seguradoras trata-se de co-seguro;
- Quando parte do risco seguro é cedida a outra seguradora em troca de prémio trata-se de resseguro.

### Características do Contrato de Seguro:



*Esquema 1.1 Características do contrato de seguro*

### **1.3.2 Ramos de Seguro**

Dada a grande diversidade de tipos de seguros, já referidos acima, existentes no mercado, houve a necessidade de classificar cada seguro com base no tipo de risco a cobrir ou no património material ou imaterial a proteger. Existem diversos ramos de seguro sendo que os que têm mais peso na carteira das companhias de seguro em termos de prémios recebidos e pagamentos efetuados são:

- Ramo automóvel;

- Ramo acidentes de trabalho;
- Ramo vida;
- Ramo saúde;
- Ramo multirriscos;
- Outros.

### *1.3.2.1 Seguro Automóvel*

O seguro automóvel é um seguro obrigatório no que respeita à cobertura de Responsabilidade Civil, e tem como finalidade proteger eventuais vítimas de acidentes de viação.

O seguro de responsabilidade civil obrigatório responde ao que é legalmente exigido relativamente à obrigação de segurar a responsabilidade civil decorrente da circulação de veículos a motor, seus reboques e semi-reboques perante:

- Terceiros não transportados: peões da via pública ou que circulem noutra viatura;
- Terceiros transportados a título oneroso: transporte coletivo de passageiros;
- Terceiros transportados a título gratuito: situações de “boleia”.

O seguro de responsabilidade civil automóvel garante a responsabilidade sobre diversos danos, entre eles:

- A responsabilidade civil pelos prejuízos causados a terceiros em virtude da circulação do veículo seguro;
- Os danos causados a terceiros, proveniente de acidentes de viação dolorosamente provocados ou resultantes de furto, roubo ou furto de uso;
- Danos nos bens transportados no veículo seguro no caso de transporte coletivo de mercadorias;
- Danos causados a terceiros derivados de acidente de viação, quando o condutor está alcoolizado, drogado ou não se encontra habilitado para a condução.

Relativamente à obrigatoriedade do seguro, todos os veículos terrestres a motor, seus reboques e semi-reboques só podem circular na via pública se tiverem seguro de responsabilidade civil, tal como qualquer prova desportiva e respetivos treinos oficiais só podem ser autorizadas mediante a existência de um seguro. No entanto existem casos em que são isentos de segurar tais como, o estado português, a menos que sejam obrigados por despacho do ministro, organizações internacionais de que o estado português seja membro e estados estrangeiros onde igual isenção seja concedida ao estado português.

O Seguro Automóvel de responsabilidade civil obrigatório cobre todo o âmbito territorial nacional, o território dos estados membros da união europeia, território de



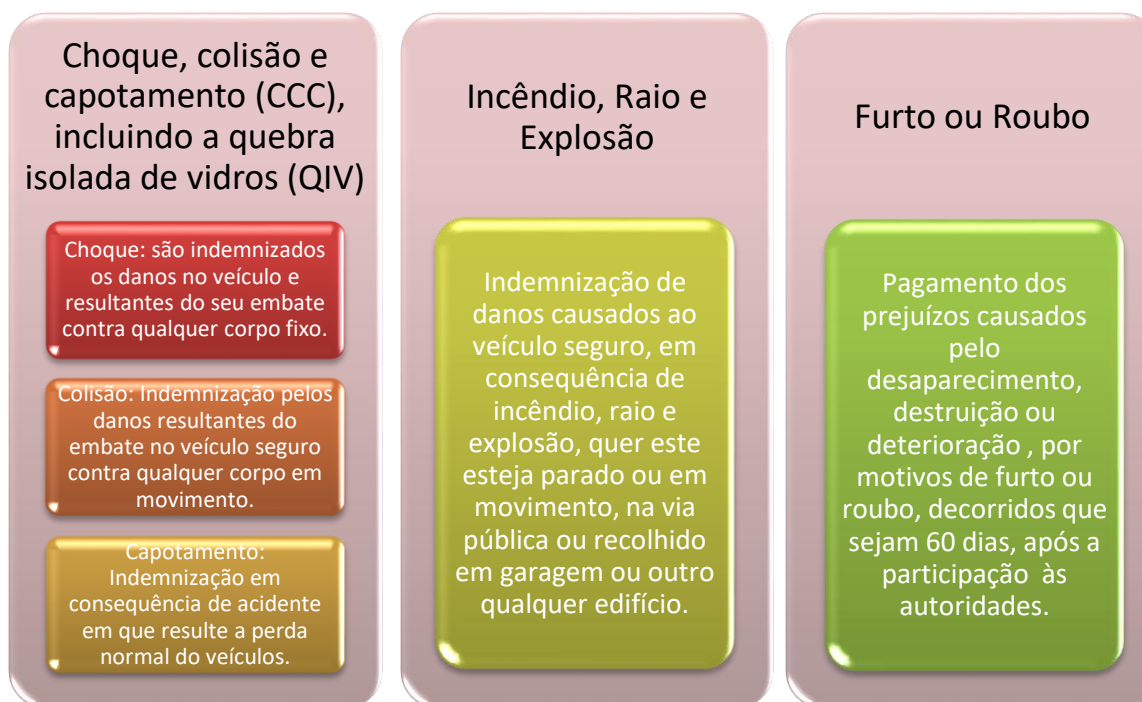
países exteriores à união europeia se forem aderentes ao acordo multilateral de garantia entre serviços nacionais de seguros ou outros aderentes indicados no certificado de seguro, assim como o trajeto que ligue diretamente dois estados membros da união europeia quando nesse território não exista serviço nacional de seguros.

Por outro lado existem exclusões que não estão abrangidas pelo seguro de responsabilidade civil obrigatórias:

- Condutor e tomador do seguro;
- Passageiros, quando transportados em contravenção do código da estrada;
- Representantes legais de pessoas coletivas ou sociedades responsáveis pelo acidente, quando no exercício das suas funções;
- Cônjuge, ascendentes ou descendentes;
- Danos causados no veículo seguro;
- Danos causados em bens transportados no veículo seguro quer durante o transporte quer durante operações de carga e descarga;
- Danos ocorridos durante a realização de qualquer prova desportiva e respetivo treino (quando não existir seguro de provas desportivas);
- Por danos originados direto ou indiretamente por explosões ou radiações atómicas.

Para cobrir algumas das exclusões acima mencionadas existe o Seguro de Danos Próprios consiste num seguro facultativo. Este tipo de seguro abrange os prejuízos sofridos pelo veículo seguro independentemente de quem é responsável pelo acidente, mesmo que seja o próprio condutor.

#### Garantias das Coberturas do Seguro de Danos Próprios:



**Figura 1.1** Garantias das Coberturas do Seguro de Danos Próprios

Resumidamente os valores a segurar nestes dois tipos de seguros já vistos são:

- **Responsabilidade Civil:** Satisfazer o mínimo imposto pelo seguro obrigatório;
- **Danos Próprios:** Valor do veículo acrescido dos extras e pintura de letras ou desenhos previstos na apólice. O valor seguro deverá corresponder ao valor venal do carro.

Ainda existe um terceiro grupo de seguros denominados de Seguros Especiais:



**Figura 1.2.** Seguros Especiais

Apenas irei especificar alguns destes oito seguros tendo em conta a maior exposição presente na carteira da seguradora AXA.

- Frotas: Este seguro destina-se a empresas que seguram simultaneamente mais de trinta veículos. Ficam excluídos do âmbito da cobertura os veículos dos trabalhadores e sócios do tomador do seguro, salvo nos casos em que, por disposição da entidade patronal seja obrigada a suportar o pagamento dos respetivos prémios.
- Seguro Automobilista: Cobre os riscos e importâncias máximas fixadas nas condições particulares quanto a sinistros ocorridos com qualquer veículo não abrangido a seguro (veículos do Estado), desde que o responsável pela condução seja portador da licença ou carta de condução.
- Provas desportivas: Terá que ser feito um seguro caso a caso que garanta a responsabilidade civil dos organizadores, proprietários dos veículos e dos seus detentores e condutores em virtude de acidentes causados pelos mesmos veículos. No entanto ficam excluídos os danos causados aos participantes e

respetivas equipas de apoio, danos causados aos veículos utilizados por participantes e equipas e os danos causados à entidade organizadora e pessoal ao seu serviço.

- **Reboques:** É realizado na mesma apólice do veículo rebocador. O seguro de tratores agrícolas, moto cultivadores e máquinas agrícolas com locomoção própria, inclui a garantia de reboque agrícola ou alfais que lhe possam ser atrelados.

### 1.3.3 Tarifa

Como já tinha sido referido anteriormente, o contrato de seguro garante a reparação ou pagamento, pela seguradora, dos danos decorrentes de um sinistro que se enquadre nas condições do referido contrato, seja o contrato de responsabilidade ou de danos próprios. A segurança/proteção dos riscos só é garantida mediante o pagamento de um prémio pelo tomador de seguro, contudo um dos principais problemas que se coloca é o do cálculo do prémio a cobrar ao tomador, que deve ser suficiente para fazer face ao custo futuro de eventuais sinistros, mas também ser suportável pelo tomador de seguro e competitivo com o mercado. Este subcapítulo começa por apresentar o conceito de prémio, bem como outros conceitos básicos subjacentes ao cálculo do prémio.

A ideia da existência de um “guia” de tarifação, que sirva de orientação ao subscritor (entidade que celebra uma operação de capitalização com o segurador, sendo responsável pelo pagamento da prestação), é seguida desde que surgiram os primeiros contratos de seguro, aplicados aos seguros marítimos, em que os prémios de seguro eram baseados nas características de cada navio a ser segurado, nomeadamente na construção e proteção de cada navio, o que resultava numa classificação dada a cada navio e que era anotada e consultada na subscrição de um novo seguro.

Hoje em dia, as tarifas têm vários fatores em consideração – os fatores de tarifação – mas o principal objetivo de uma tarifa continua a ser o de construir uma medida adequada do risco em que a seguradora incorre ao aceitar subscrever uma determinada apólice. No entanto, a construção de uma tarifa deve ter em conta dois aspetos fundamentais:

- A adequabilidade do prémio global;
- A alocação correta do prémio a cada fator de tarifação.

O volume global de prémios deve ser adequado, de forma a manter o equilíbrio técnico, ou seja, deve ser suficiente para garantir o pagamento das responsabilidades futuras, das despesas e para fazer face ao desenvolvimento adverso da sinistralidade. Por outro lado, a aplicação de um prémio igual a todos os riscos não é desejável, pois tal poderia originar uma situação conhecida como “anti seleção”, ou seja, numa situação em que todos os fatores de risco pagam o mesmo prémio, a companhia poderá “atrair” os maus

riscos. A correta alocação do prêmio a cada fator de tarifação pode ainda permitir melhorar a competitividade e rentabilidade da seguradora no mercado.

O projeto debruça-se principalmente sobre um dos fatores de tarifação – zona tarifária – através de uma análise estatística dos sinistros automóveis ocorridos em 2013, 2014 e 2015, usando um dos principais indicadores de sinistralidade não vida (frequência).

Definição de alguns conceitos essenciais à modelação realizada:

- Prémio Total: contrapartida da cobertura acordada, incluindo o conjunto dos custos que impedem sobre o tomador do seguro, nomeadamente custos de cobertura de risco, custos de aquisição, de gestão e de cobrança e encargos relacionados com a emissão da apólice.
- Prémio Puro: valor que custa ter um conjunto de apólices em vigor, ou seja, se cobrarmos exatamente este prêmio puro por cada apólice apenas se consegue pagar os custos relativos a essas apólices e nada mais. Também pode ser designado por custo por unidade de tempo.
- Prémio Comercial: Entre o prêmio de risco (puro) e o prêmio comercial pode haver uma grande diferença, porque ao prêmio puro é preciso acrescentar outros encargos necessários para o negócio se manter em funcionamento. Estes encargos são de diversa espécie, fixos e variáveis, tais como custo de exploração e manutenção, custo de reinvestimento, margem de lucro, margem de segurança, inflação, comissões, resseguro, impostos...
- Sobreprémio: Majoração ou suplemento de prêmio que corresponde à cobertura de um risco mais grave que o risco normal, ou a uma garantia suplementar.
- Bonificação: Redução do prêmio de renovação do contrato de seguro, verificado que houve ausência de sinistros no período de contrato antes da renovação.
- Agravamento: Aumento do prêmio de renovação do contrato de seguro, verificado que houve ocorrência de sinistros no período de contrato antes da renovação.
- Sistema Bónus/Malus: Consiste numa escala progressiva de agravamentos por sinistros e de bónus por ausência do mesmo. De igual modo a aplicação do primeiro agravamento não implica a retirada do bónus anteriormente atribuído.
- Exposição: Conjunto de contratos expostos ao risco que constituem a carteira de uma seguradora.
- Frequência: Proporção de sinistros por exposição ao risco.
- Zona tarifária: Corresponde à zona de circulação do veículo, a qual é determinada pelo concelho de residência do segurado.
- Segmentação: É um conceito de mercado muito importante e em termos de risco faz todo o sentido, porque uma tarifa corretamente segmentada permite reter clientes e “taxar” convenientemente os maus riscos.
- Ecretagem: A existência de sinistros particularmente elevados “transtorna” os níveis de sinistralidade de determinados segmentos, o que pode dar a ideia que têm características “más” o que pode não corresponder à verdade. A ecretagem é um “corte” no valor dos sinistros para que todo e qualquer sinistro tenha no

máximo um valor arbitrário denominado nível de ecretagem. Assim a sinistralidade global é mais baixa e é alisado o efeito dos sinistros muito elevados.

- Sinistros IBNR: Sinistros ocorridos mas não reportados (incurred but not reported). Este tipo de sinistros são muito importante pois acarretam um conjunto de custos associado ao risco atual mas que ainda não temos conhecimento.
- Sinistralidade: Número de sinistros.
- Loss-ratio: Proporção do custo do sinistro por prémio adquirido.

## Capítulo 2. Enquadramento Teórico

### 2.1 Frequência: Conceito

Na análise e gestão de carteira de seguros é necessário construir um conjunto de indicadores que nos permitam de forma resumida obter uma ideia de como a carteira se está a comportar, para tal é útil saber:

- o nível de prémio;
- intensidade;
- anulações;
- sinistralidade;
- frequência.

Neste subcapítulo explorar-se-ão os últimos dois pontos.

Intuitivamente, a frequência de sinistralidade diz-nos com que relativa assiduidade ocorrem sinistros. Esta medida representa o número de sinistros que ocorrem num determinado conjunto de apólices sendo este um **indicador de sinistralidade**. Contudo a definição mais correta será que a **frequência** é uma medida relativa do número de sinistros ou proporção de sinistros por exposição ao risco e é calculado da seguinte forma:

$$Frequência = \frac{N^{\circ} \text{ de sinistros}}{Exposição \text{ ao risco}}$$

*Equação 2.1 Frequência de Sinistros*

No mínimo este indicador é zero quando o nº de sinistros é zero e não existe valor máximo. Então podemos concluir que quanto menor for este indicador melhor é o cliente.

O número de sinistros é fácil de decifrar, porque tal como foi dito anteriormente este valor indica a quantidade de ocorrências de um acontecimento incerto previsto no contrato, mas o que é exposição ao risco?

A seguradora só está exposta ao risco do conjunto de contratos que constituem a sua carteira após a data de início do contrato até à sua anulação. Este intervalo temporal é o objetivo de estudo da exposição ao risco.

A exposição é um conceito importante em análise pois é uma medida “exata” dos níveis de risco. Usualmente o nível de exposição é medido em “anos seguros”, ou seja, se um seguro estiver válido durante um ano inteiro terá exposição 1. Se o seguro apenas vigorar durante meio ano (por exemplo anulou após 6 meses da data de início ou apenas fez o seguro há 6 meses) este seguro terá exposição 0.5. Assim esta é uma medida da proporção do tempo de exposição ao risco e ajuda a perceber as características de uma determinada carteira.

Em termos de análise técnica de uma carteira a exposição é o tempo exposto de cada um dos contratos dentro de um intervalo de tempo fixo.

### **Exemplo prático:**

Intervalo de tempo de estudo: 01/01/2015 a 31/12/2015 (12 meses - 365 dias)

- O contrato A tem início a 01/01/2015 e anulou a 03/03/2015;
- O contrato B tem início a 01/01/2014 e anulou a 03/03/2015;
- O contrato C tem início a 01/12/2014 e anulou a 30/06/2015;
- O contrato D tem início a 01/04/2015 e anulou a 30/06/2015;
- O contrato E tem início a 01/04/2015 e continua em vigor.

*Tabela 2.1. Exemplo Prático*

Contrato	Dias em vigor no ano 2015	Exposição ao risco
A	62	$\frac{62}{365}$
B	62	$\frac{62}{365}$
C	181	$\frac{181}{365}$
D	91	$\frac{91}{365}$
E	275	$\frac{275}{365}$

Pelo que a exposição total desta carteira é:

$$\frac{62}{365}(A) + \frac{62}{365}(B) + \frac{181}{365}(C) + \frac{91}{365}(D) + \frac{275}{365}(E)$$

*Equação 2.2. Cálculo da exposição*

Agora suponhamos neste mesmo intervalo de tempo o número de sinistros por cada contrato:

- O contrato A teve 3 sinistros em 2015;
- O contrato B teve 1 sinistro em 2014 e outro em 2015;
- O contrato C nunca teve nenhum sinistro;
- O contrato D teve 1 sinistro a 01/04/2015 e outro em 09/06/2015;
- O contrato E teve 1 sinistro em 01/01/2016.

**Tabela 2.2.** Resultados do exemplo prático

Contrato	#Sinistros no ano 2015	Exposição ao risco	Frequência
A	3	$\frac{62}{365}$	$\frac{\frac{3}{62}}{\frac{1}{365}} = \frac{1095}{62} \approx 17.66$
B	1	$\frac{62}{365}$	$\frac{\frac{1}{62}}{\frac{1}{365}} = \frac{365}{62} \approx 5.888$
C	0	$\frac{181}{365}$	0
D	2	$\frac{91}{365}$	$\frac{\frac{2}{91}}{\frac{1}{365}} = \frac{730}{91} \approx 8.022$
E	0	$\frac{275}{365}$	0

Se quiséssemos calcular a frequência da carteira total (dos 5 contratos em conjunto) teríamos que somar os sinistros e a exposição ao risco de todos os contratos, e efetuar a respetiva divisão.

## 2.2 Modelo Linear Generalizado

### 2.2.1 Definição

Hoje em dia, os modelos lineares generalizados, também conhecidos por Generalized Linear Models (GLM), são amplamente reconhecidos como o método padrão da indústria para a tarifação do ramo automóvel tanto na linha individual como em pequenos ramos da linha comercial. A grande parte dos países europeus, por exemplo, Reino Unido, Irlanda, França e mesmo Portugal, usam os GLM's para analisar os seus portfólios do ramo automóvel. Este método está a ganhar popularidade em países da Europa Leste e da Ásia porque não só é tecnicamente eficiente como permite explicitar a sua relação da tarifa com variáveis preditivas. Além disso os GLM's fornecem diagnósticos estatísticos que ajudam a selecionar apenas as variáveis significativas para o modelo.

As principais aplicações do modelo linear generalizado na análise da carteira de uma seguradora são:

- Na tarifação;
- E na subscrição.

A procura de métodos eficientes que diminuam o número e custo (por exemplo implementação do IT) associado à mudança das taxas de tarifação, fez aumentar cada vez mais o uso de modelos lineares generalizados para uma análise do mercado-alvo.



No passado, os atuários e analistas dependiam fortemente da análise “one-way” para analisar o preço e o monitorizar a performance da tarifa no mercado. Esta análise “one-way” resume indicadores de sinistralidade (estatísticas), tais como a frequência e o loss-ratio, para cada valor de cada variável explicativa contudo sem contar com o efeito das outras variáveis. Pelo que este tipo de análise pode ser distorcido pelas correlações entre fatores. Por exemplo, condutores jovens normalmente conduzem carros velhos, a análise “one-way” com idade do carro pode demonstrar um número elevado de sinistros para carros mais velhos, no entanto isto pode resultar apenas do facto que este tipo de carro ser geralmente conduzido por condutores jovens que representam um elevado risco, porque por norma conduzem há menos tempo. As técnicas atuariais tradicionais para lidar com este problema tentam padronizar os dados, de modo a remover este efeito de distorção, porque este método é apenas uma aproximação da realidade. Métodos de regressão múltipla como o modelo linear generalizado ajustam correlações e permitem o estudo sobre a interação entre variáveis, ou seja, analisam a influência que uma ou mais variáveis explicativas, medidas em indivíduos ou objetos, têm sobre uma variável de interesse (variável resposta).

Os Modelos Lineares Generalizados foram introduzidos por Nelder e Wedderburn em 1972 e correspondem a uma síntese de vários modelos até então desenvolvidos, vindo assim unificar tanto do ponto de vista teórica como concetual, a teoria da modelação estatística.

O objetivo do modelo linear generalizado é expressar a relação entre uma variável aleatória  $Y$  de interesse primário, que denominamos de variável resposta ou variável dependente, e um vetor  $\mathbf{X} = (x_1, \dots, x_k)^T$  de  $k$  variáveis explicativas, também designadas por covariáveis ou variáveis independentes que explicam parte da variabilidade inerente a  $Y$ . A variável  $Y$  pode ser:

- Contínua: corresponde a características mensuráveis que assumem valores numa escala contínua, ou seja, a variável pode assumir teoricamente qualquer valor num intervalo.
- Discreta: corresponde a características mensuráveis que só podem assumir valores inteiros.
- Dicotómica: corresponde a variáveis qualitativas em que só há duas respostas possíveis do tipo Sim/Não, Doente/ Não Doente, óbito/vivo.

As covariáveis podem ser também de qualquer natureza: contínuas, discretas, dicotómicas ou qualitativas de natureza ordinal (qualitativas porque são características que se baseiam em qualidade e não podem ser mensuradas numericamente, e são de natureza ordinal porque existe uma ordenação entre as categorias).

Os GLM apresentam vantagens por serem uma generalização dos modelos lineares, entre as quais, a flexibilidade da função de regressão. A relação entre a variável resposta e a combinação linear das variáveis explicativas é habitualmente efetuada por uma

função  $\theta$  denominada de função de ligação. A maior vantagem desta função é a possibilidade de construção de modelos que permitem obter intervalos de confiança das estimativas, com base em distribuições pertencentes à família exponencial.

### 2.2.2 Família Exponencial

Assumimos que temos os dados da forma:  $(Y_i, \mathbf{X}_i), i = 1, \dots, n$ , resultantes da realização de  $(Y, \mathbf{X})$ , em  $n$  indivíduos, sendo as componentes  $Y_i$  do vetor aleatório  $Y = (Y_1, \dots, Y_n)^T$  independentes. Na forma matricial representamos as variáveis da seguinte forma:

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

Em muitas situações práticas, especialmente quando as variáveis explicativas são de natureza qualitativa, existem muitos indivíduos na amostra que partilham do mesmo vetor de covariáveis, pelo que é mais prático e perceptível apresentar os dados de uma forma agrupada. Podemos associar os indivíduos em  $g$  grupos distintos, de tal modo que os  $n_j$  indivíduos do grupo  $j$  ( $j = 1, \dots, g$  com  $\sum_{j=1}^g n_j = n$ ) partilhem do mesmo vetor de covariáveis, digamos  $\mathbf{x}_j = (x_{j1}, \dots, x_{jk})^T$ . Os dados passarão a ser então representados por:

$$\bar{\mathbf{y}} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_g \end{pmatrix} \quad X_g = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{g1} & \cdots & x_{gk} \end{pmatrix}$$

Onde  $\bar{y}_j, j = 1, \dots, g$ , representa a média das variáveis respostas dos indivíduos que pertencem ao  $j$ -ésimo grupo e não existem linhas idênticas em  $X_g$ . O agrupamento dos dados é particularmente importante em situações em que as covariáveis são todas de natureza qualitativa.

Os modelos lineares generalizados pressupõem que a variável resposta tenha uma distribuição pertencente à família exponencial. Diz-se que a variável  $Y$  tem distribuição pertencente à família exponencial se a sua função densidade de probabilidade ou função massa de probabilidade se possa escrever na forma:

$$f(y|\theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\},$$

- $\theta$ : é a forma canónica do parâmetro escalar de localização;
- $\phi$ : é um parâmetro escalar de dispersão conhecido;
- $a(\cdot), b(\cdot)$  e  $c(\cdot, \cdot)$ : são funções reais conhecidas, sendo que a distribuição de  $b(\cdot)$  não depende dos parâmetros e  $a(\phi) = \frac{\phi}{\omega}$ , onde  $\omega$  é uma constante conhecida.

Passarei agora para o valor médio e a variância.

Seja  $\ell(\theta; \phi, y) = \ln(f(y|\theta, \phi))$  e a função score  $S(\theta) = \frac{\partial \ell(\theta; \phi, Y)}{\partial \theta}$  então:

- $E(S(\theta)) = 0$
- $E(S^2(\theta)) = E\left[\left(\frac{\partial \ell(\theta; \phi, Y)}{\partial \theta}\right)^2\right] = -E\left[\left(\frac{\partial^2 \ell(\theta; \phi, Y)}{\partial \theta^2}\right)\right]$

Como  $\ell(\theta; \phi, y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$  obtêm-se  $S(\theta) = \frac{Y - b'(\theta)}{a(\phi)}$  e  $\frac{\partial S(\theta)}{\partial \theta} = -\frac{b''(\theta)}{a(\phi)}$ , onde  $b'(\theta) = \frac{\partial b(\theta)}{\partial \theta}$  e  $b''(\theta) = \frac{\partial^2 b(\theta)}{\partial \theta^2}$ .

Assim:

- $E(Y) = \mu = a(\phi)E(S(\theta)) + b'(\theta) = b'(\theta) \rightarrow$  Valor médio
- $Var(Y) = a^2(\phi)Var(S(\theta)) = a^2(\phi)\frac{b''(\theta)}{a(\phi)} = a(\phi)b''(\theta) \rightarrow$  Variância

Algumas distribuições da família exponencial:

**Tabela 2.3.** Distribuições da família exponencial

Distribuição	Normal	Binomial	Poisson	Gama
Notação	$N(\mu, \sigma^2)$	$B(m, \pi)/m$	$P(\lambda)$	$Ga(v, \frac{v}{\mu})$
Suporte	$(-\infty, +\infty)$	$\{0, \frac{1}{m}, \dots, 1\}$	$\{0, \dots, 1\}$	$(0, +\infty)$
$\theta$	$\mu$	$\ln\left(\frac{\pi}{1-\pi}\right)$	$\ln(\lambda)$	$-\frac{1}{\mu}$
$a(\phi)$	$\sigma^2$	$\frac{1}{m}$	1	$\frac{1}{v}$
$\phi$	$\sigma^2$	1	1	$\frac{1}{v}$
$\omega$	1	$m$	1	1
$c(y, \phi)$	$-\frac{1}{2}\left(\frac{y^2}{\phi} + \ln(2\pi\phi)\right)$	$\ln\left(\frac{m}{my}\right)$	$-\ln(\lambda)$	$v \ln v - \ln \Gamma(v) + (v-1) \ln(y)$
$b(\theta)$	$\frac{\theta^2}{2}$	$\ln(1 + e^\theta)$	$e^\theta$	$-\ln(-\theta)$
$b'(\theta) = E(Y)$	$\theta$	$\pi = \frac{e^\theta}{1 + e^\theta}$	$\lambda = e^\theta$	$\mu = \frac{1}{\theta}$
$b''(\theta) = V(\mu)$	1	$\pi(1 - \pi)$	$\lambda$	$\mu^2$
Var(Y)	$\sigma^2$	$\frac{\pi(1 - \pi)}{m}$	$\lambda$	$\frac{\mu^2}{v}$

Tal como já foi referenciado os modelos lineares generalizados têm uma grande importância do mercado segurador, o seguinte quadro resume alguns dos modelos mais utilizados na atividade seguradora:

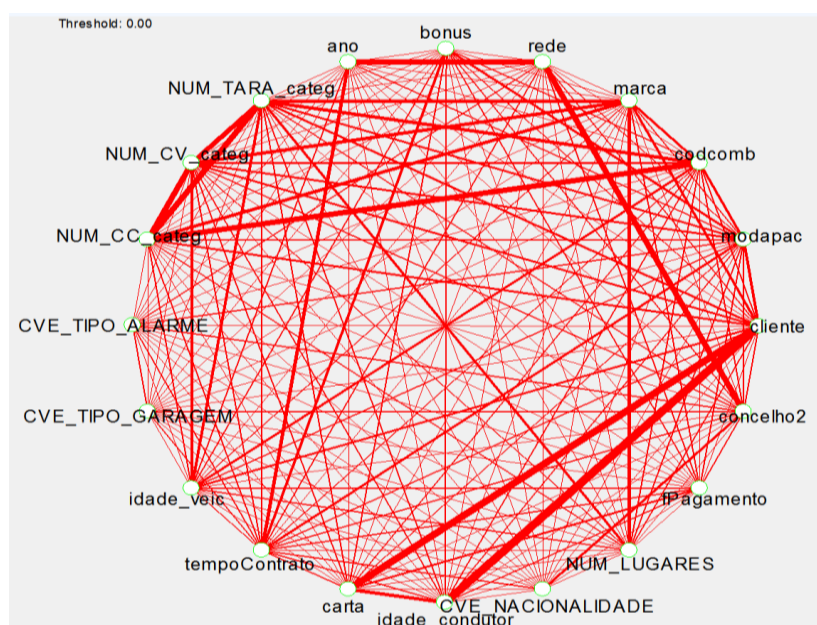
**Tabela 2.4. Modelo Lineares Generalizados**

$Y$	Frequência de sinistros	Quantidade de sinistros	Custo médio	Renovações
<b>Função</b>	$\ln(x)$	$\ln(x)$	$\ln(x)$	$\ln\left(\frac{x}{1-x}\right)$
<b>Distribuição</b>	Poisson	Poisson	Gamma	Binomial
$\phi$	1	1	Estimado	1
$V(x)$	$x$	$x$	$x^2$	$x(1-x)$
$\omega$	Exposição	1	Nº de sinistros	1

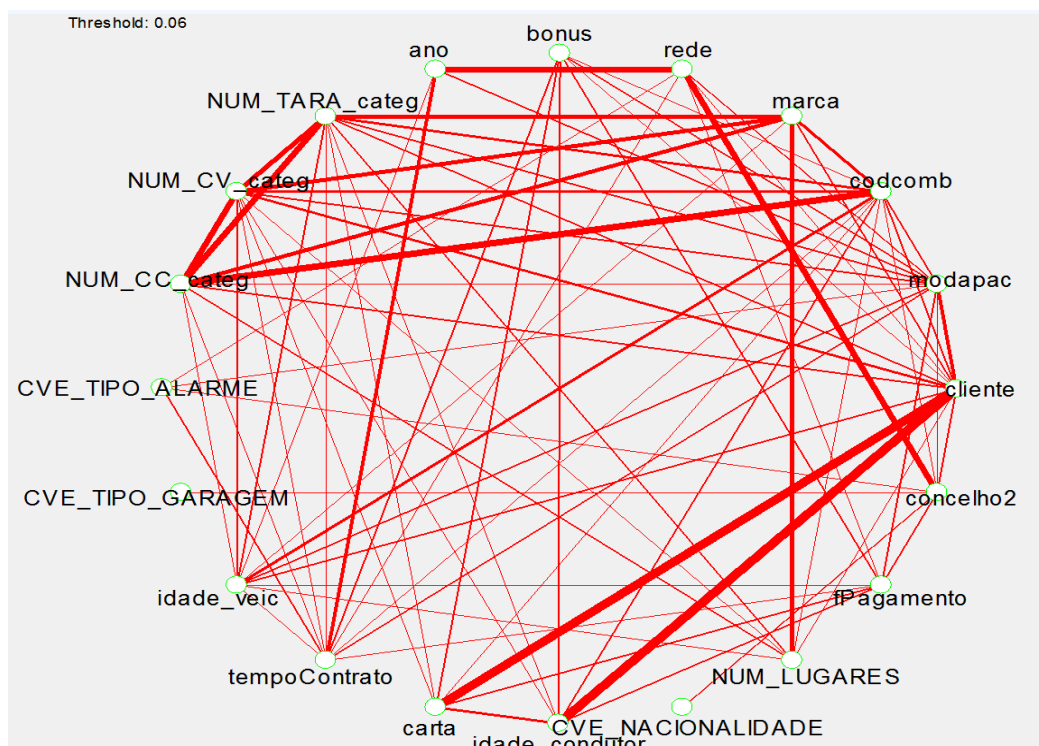
Para mais pormenores consultar o livro “*Modelos Lineares Generalizados – da Teoria à Prática*”(2000).

### 2.3 Métodos de Correlação

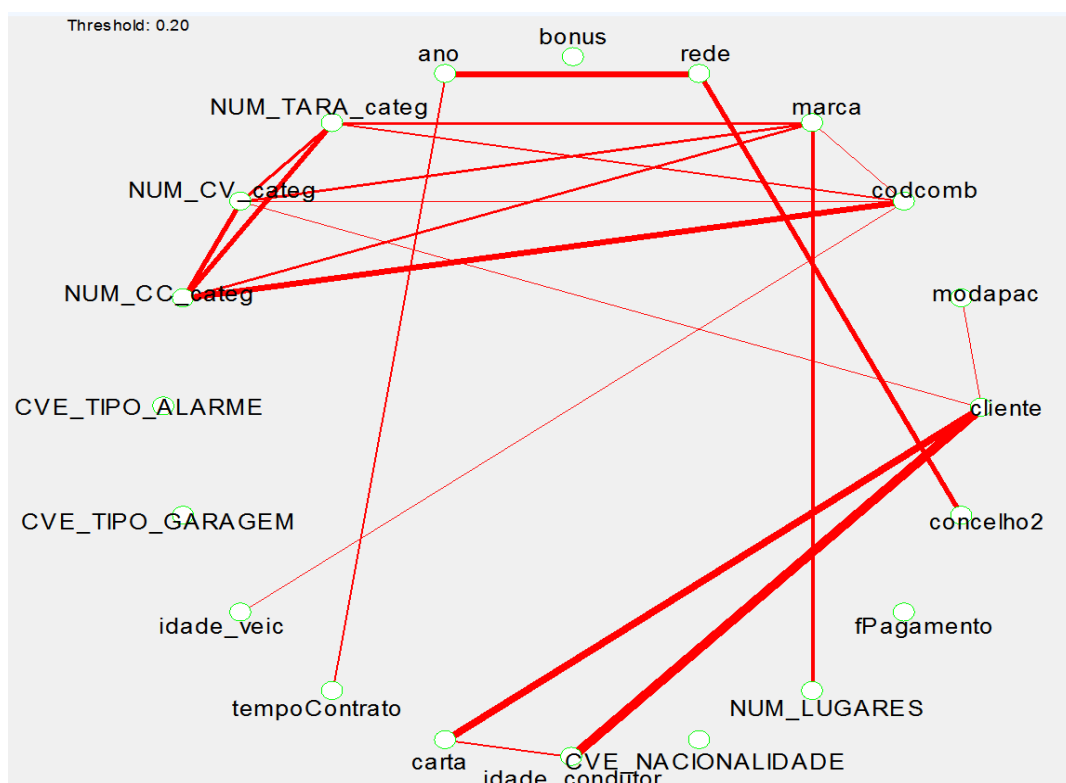
Um dos métodos mais diretos de reduzir o número de variáveis numa primeira interpelação ao modelo, é a análise da correlação das variáveis. É preciso ter cuidado para não incluir duas ou mais variáveis fortemente correlacionadas entre si, pois o objetivo é determinar a verdadeira contribuição de cada uma das variáveis para a variável resposta. Se as variáveis estiverem muito relacionadas entre si, pequenas mudanças nos dados podem originar grandes variações dos coeficientes resultando mesmo em erros no modelo final. A este fenómeno a Estatística denomina de multicolinearidade. Neste projeto foi estabelecido um limite de correlação de 0.7, valor usual na atividade seguradora, sendo que todas as variáveis que reflitam informações duplicadas são eliminadas.



**Figura 2.1. Correlações entre as variáveis sem qualquer limite**



**Figura 2.2.** Correlações entre variáveis acima de 0.06



**Figura 2.3.** Correlações entre variáveis acima de 0.2

## 2.4 Seleção de Variáveis

A abordagem tradicional na construção de modelos estatísticos é encontrar o modelo mais simples que explica os dados, ou seja, um modelo parcimonioso. Existem algumas técnicas para auxiliar na seleção de variáveis. Um dos métodos mais usados é o método Stepwise.

Construir um modelo que inclua apenas um subconjunto de variáveis explicativas envolve dois objetivos conflituosos:

1. Obter o máximo de informação por meio de um modelo com tantas variáveis independentes possíveis;
2. Diminuir o número possível de variáveis explicativas para aumentar a simplicidade do modelo.

Qualquer procedimento para a seleção ou exclusão de variáveis de um modelo é baseado num algoritmo que verifica a importância das variáveis, incluindo ou excluindo-as do modelo tendo em conta uma regra de decisão. A importância da variável é definida em termos de uma medida de significância estatística do coeficiente associado à variável para o modelo, sendo que essa estatística depende das suposições do modelo (para mais pormenores consultar o artigo “*Model selection and estimation in regression with grouped variables*”).

O critério para a adição ou remoção de variáveis é geralmente baseado na estatística F, comparando modelos com e sem as variáveis em questão. Existem três procedimentos automáticos:

- Método *Forward*;
- Método *Backward*;
- Método *Stepwise*.

### 2.4.1 Métodos de seleção

Antes da esquematização dos métodos primeiro têm de ser definidas algumas medidas:

- a)  **$SQ_{tot}$** : Soma Total dos Quadrados, ou seja, soma dos quadrados das diferenças entre a média e cada valor observado. Representa-se por:

$\sum_{i=1}^n (y_i - \bar{y})^2$ , onde  $n$  é o número de observações,  $y_i$  o valor observado e  $\bar{y}$  a médias das observações.

- b)  **$SQ_e$** : Soma dos Quadrados Explicada, indica a diferença entre a média das observações e o valor estimado para cada observação. Quanto menor for a diferença, maior poder explicativo detém o modelo. Representa-se por:

$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ , onde  $\hat{y}_i$  é o valor estimado de  $y_i$ .

- c)  **$SQ_r$** : Soma dos Quadrados dos Resíduos, que calcula a parte que não é explicada pelo modelo. Representa-se por:

$\sum_{i=1}^n (\hat{y}_i - y_i)^2$ .

Pelo que podemos determinar que:  $SQ_{tot} = SQ_e + SQ_r$

- d)  **$R^2$** : Coeficiente de determinação, é uma medida de ajustamento do modelo linear generalizado. O  $R^2$  varia entre 0 e 1 e indica o quanto o modelo consegue explicar os valores observados. Quanto maior o  $R^2$ , mais explicativo é o modelo e melhor ele se ajusta à amostra. Representa-se por:

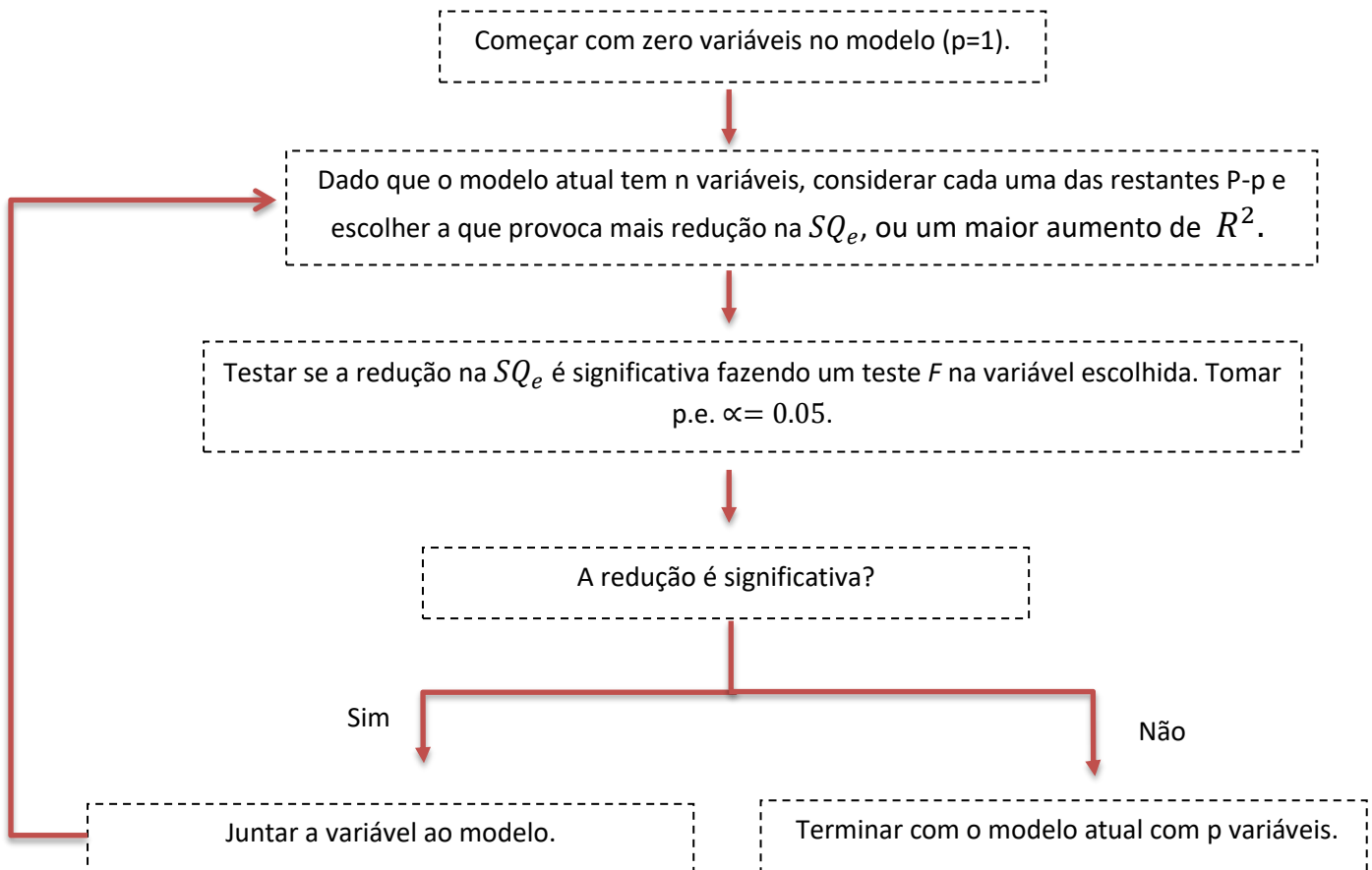
$$R^2 = \frac{SQ_e}{SQ_{tot}} = 1 - \frac{SQ_r}{SQ_{tot}}, \text{ com } 0 \leq R^2 \leq 1.$$

*Equação 2.3 Coeficiente  $R^2$*

### Método Forward

Este procedimento parte da suposição de que não existe nenhuma variável no modelo. Sendo que o princípio deste método é adicionar uma variável de cada vez até que já não exista nenhuma variável que melhore e explique o modelo.

Esquema do modelo:



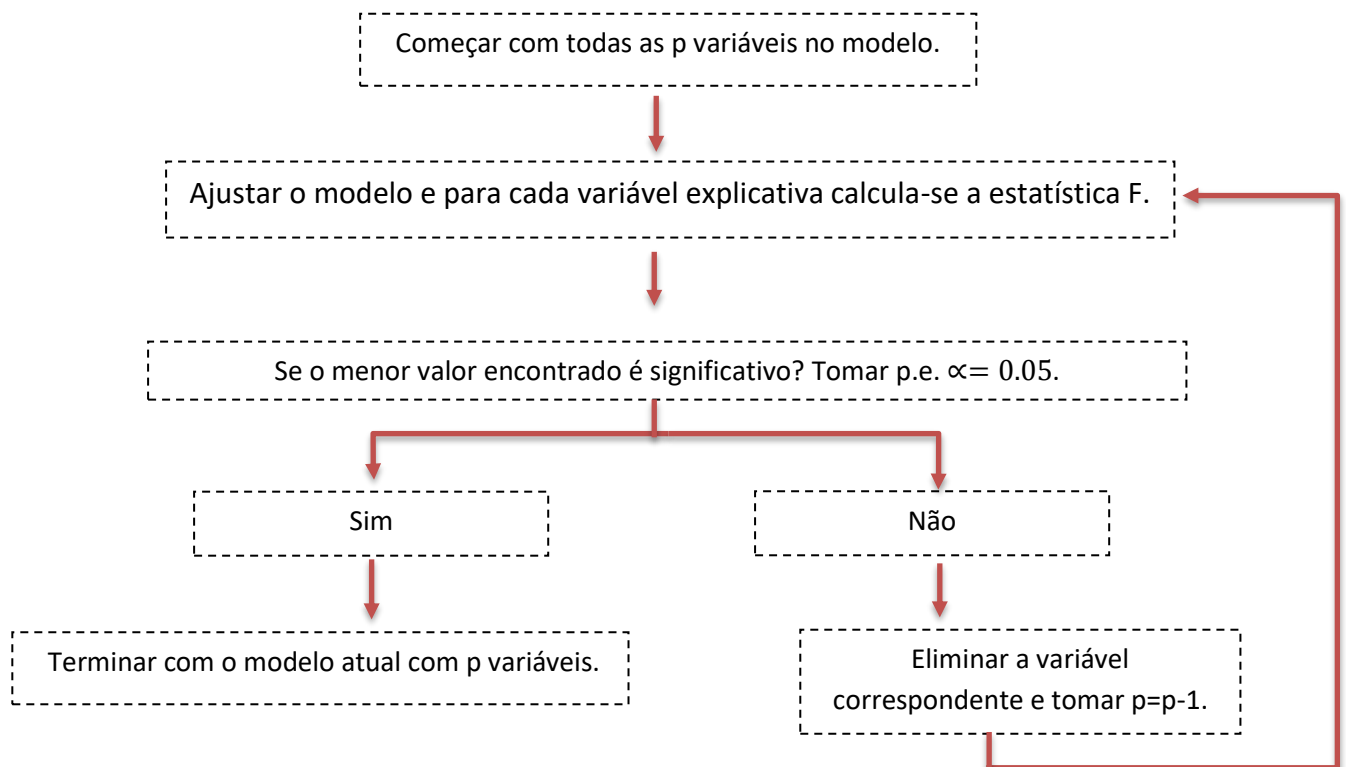
**Esquema 2.1.** Método Forward



### Método Backward

Enquanto o método *Forward* começa sem variáveis no modelo e adiciona variáveis a cada passo, o Método *Backward* faz exatamente o oposto. O processo incorpora inicialmente todas as variáveis e depois, por etapas, cada uma das variáveis pode ou não ser eliminada. A eliminação de uma variável baseia-se em testes F, que são calculados para cada variável como se a variável fosse a última a entrar no modelo.

Esquema do modelo:

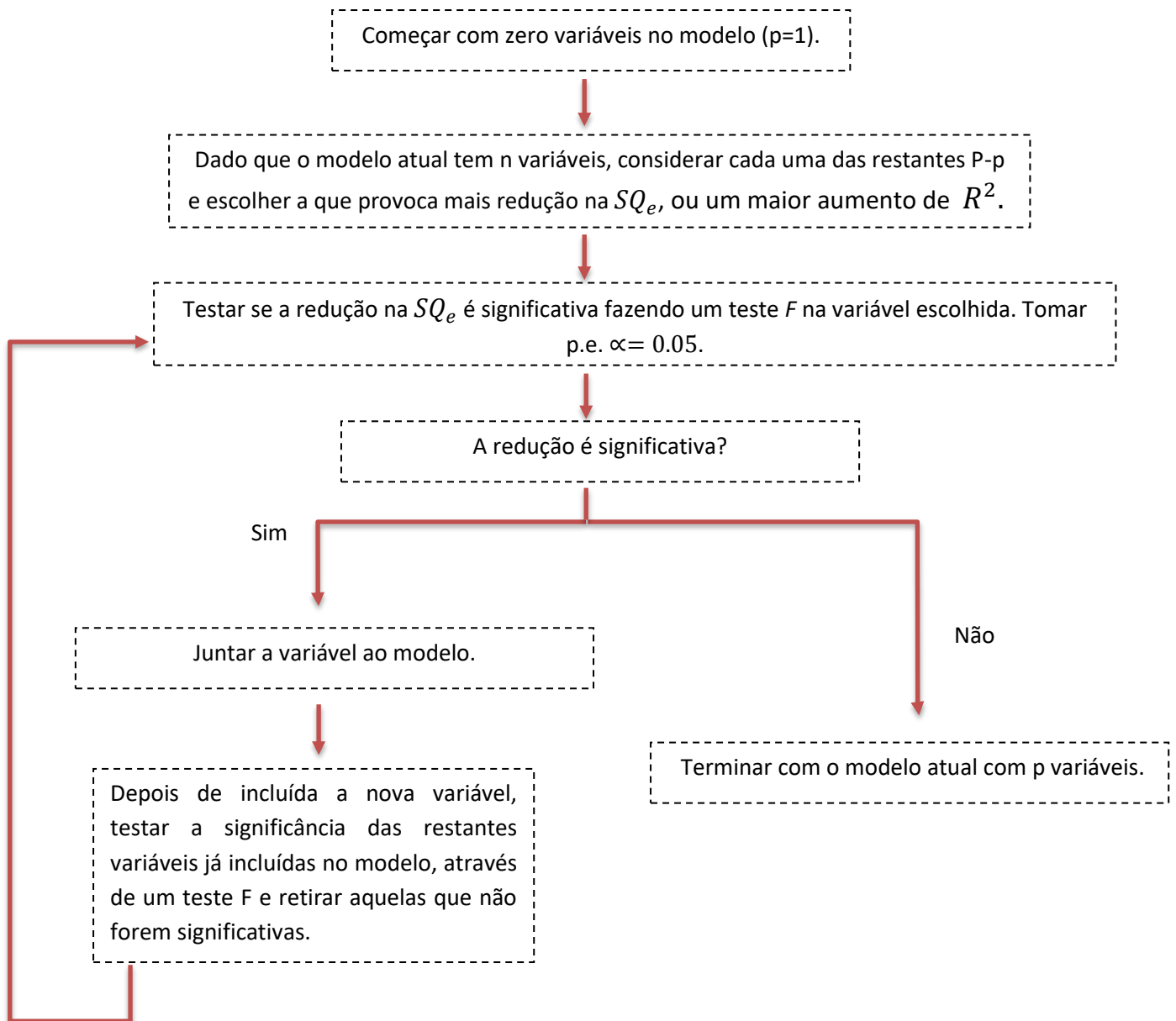


**Esquema 2.2.** Método Backward

### Método Stepwise

Esta técnica começa sem quaisquer variáveis no modelo. Para cada uma das variáveis candidatas ao modelo, o método calcula estatísticas  $F$  que refletem a contribuição que a variável traria para o modelo se fosse usada. Uma vez acrescentada uma variável, o método *Stepwise* avalia as variáveis presentes no modelo e remove quaisquer que não atinjam o critério de corte. O *Stepwise* é uma combinação dos dois métodos em cima referidos, em cada passo as variáveis são testadas para ser incluídas ou eliminadas do modelo.

Esquema do modelo:



**Esquema 2.3.** Método Stepwise

### 2.4.2 AICc

A escolha do modelo apropriado do ponto de vista estatístico é um tópico extremamente importante na análise de dados, isto porque se procura o modelo que envolva o menor número de variáveis possíveis e que expliquem bem o comportamento da variável resposta. Deste modo, é necessário fazer a seleção do melhor modelo, de entre aqueles que foram ajustados, para explicar o fenómeno sob estudo.

Os métodos de seleção podem-se basear em diversos critérios de seleção sendo um dos mais utilizados o Critério de Informação de Akaike corrigido.

Em 1973 Akaike utilizou a Informação de K-L (Kullback-Leibler) baseada na função de Máxima Verossimilhança para testar se um dado modelo é adequado. Contudo em 1987 Bozdogan propôs uma correção para o AIC. O critério de Informação de Akaike corrigido (AICc) admite a existência de um modelo “real” que é desconhecido que descreve os dados, e testa escolher de entre um grupo de modelos avaliados, o que minimiza a divergência de Kullback-Leibler. Esta divergência está relacionada com o fato que se usar um modelo aproximado e não o “real”. Por isso o modelo com menor valor AIC é considerado o melhor modelo.

O critério de Informação de Akaike é definido como:

$$AICc = -2 \log(L_p) + 2p + 2 \frac{p(p+1)}{n-p-1}$$

*Equação 2.4. Cálculo do AICc*

Em que:

- $L_p$ : é a função de máxima verossimilhança do modelo;
- $p$  é o número de variáveis explicativas consideradas no modelo;
- $n$  é o número de observações na amostra.

## Capítulo 3. Caso de Estudo: Portugal Continental

### 3.1 Preparação dos dados

Para este projeto os dados utilizados em termos temporais são de 2013 a 2015 (anos completos). Foram utilizados 3 anos porque em primeiro lugar para termos um bom modelo tem de existir uma boa base de dados por trás, para podermos modelar e melhor prever eventos futuros, e em segundo lugar existem concelhos com pouca população e como tal não terão tantos seguros de automóveis ligeiros, e sendo este um projeto para a modelação das zonas tarifárias convém ter uma vasta informação para não incorrer em erro.

Em termos de informação esta base de dados apenas inclui apólices de veículos ligeiros com Responsabilidade Civil, não só porque o veículo ligeiro é um dos meios de transporte pessoal mais usados como também o seguro de responsabilidade civil é obrigatório para todos os carros.

De acordo com os dados que a AXA recolhe dos seus clientes, temos as seguintes possíveis variáveis explicativas para descrever a frequência de sinistralidade:

*Tabela 3.1. Descrição das variáveis*

Variável	Descrição
→ Bonus	Escalão de Bónus-Malus
→ zona	Classificação da rede de agentes
→ marca	Marca do veículo
→ codcomb	Tipo de combustível
→ pack	Pack Auto
→ cliente	Tipo de cliente (Empresa/Particular)
→ concelho2	Concelho
→ fPagamento	Forma de Pagamento
→ lugares	Número de lugares do veículo
→ CVE_NACIONALIDADE	Nacionalidade do Tomador
→ idadecond	Idade do Tomador
→ idadecart	Tempo de carta do Tomador
→ contrato	Tempo de contrato da apólice
→ idadeveic	Idade do veículo
→ CVE_TIPO_GARAGEM	Código de existência ou não de garagem e o tipo (noite/dia)
→ CVE_TIPO_ALARME	Tipo de Alarme
→ NUM_CC_categ	Cilindrada do veículo
→ NUM_CV_categ	Potência do veículo
→ NUM_TARA_categ	Tara do veículo
→ ano	Ano em questão

Poderiam ter sido incluídas outras variáveis para o projeto, contudo infelizmente alguma informação não é 100% fiável e como alguns campos não são de preenchimento obrigatório para a realização do contrato de seguro não são preenchidos e por isso algumas variáveis podem ter pouca ou mesmo nenhuma informação. O género não foi incluído como variável explicativa porque não é aconselhável a criação de uma tarifa em que uma das variáveis de risco seja o sexo do tomador ou do condutor habitual do veículo, e como este projeto é uma projeção da realidade pelo que optou-se por não se incluir a variável.

## 3.2 Frequência

O primeiro passo para a modelação da base de estudo é analisar a frequência real sem qualquer tipo de modelo embutido, nem analisando nenhum tipo de fator de risco específico. Nesta fase foi calculada a frequência de todos os concelhos de Portugal Continental, pois o objetivo deste projeto é modelar as zonas tarifárias. Como tal depois de calcular a frequência real, foram efetuados dois tipos de análises. A estas duas análises denominei por:

- Frequência Ponderada;
- Frequência Conjunto.

### 3.2.1 Frequência Real

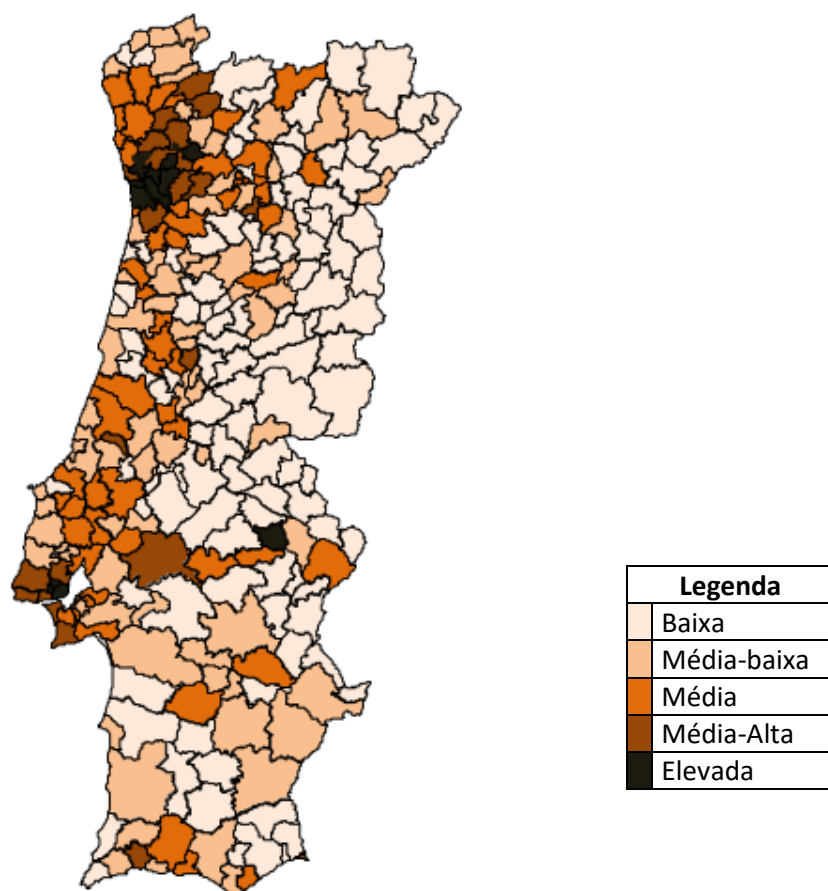
Esta foi a primeira análise do projeto. Calculei a exposição e o número de sinistros de todos os clientes, agregando-os por concelhos de Portugal Continental, com base num período de 3 anos (2013-2015). Desta forma foi possível obter uma visão completa da frequência de sinistralidade dos clientes.

Para a conceção de todos os mapas quis criar uma legenda única com a forma da distribuição Normal para poder ter uma melhor e correta leitura dos gráficos. Tive em consideração todos os dados dos 3 anos da base de estudo, em que calculei a frequência por concelho e agrupei os concelhos em 5 zonas tarifárias tendo em conta o valor da frequência e os respetivos quantis. Os valores definidos para o intervalo de frequência obtiveram-se de modo empírico. Pelo que a legenda ficou da seguinte forma:

*Tabela 3.2. Legenda da Frequência*

Legenda da Frequência		Intervalo de frequência
	Baixa	$(0, X_{(0,0.17)})$
	Média-baixa	$(X_{(0,0.17)}, (X_{(0,0.37)})$
	Média	$(X_{(0,0.37)}, (X_{(0,0.62)})$
	Média-Alta	$(X_{(0,0.62)}, (X_{(0,0.82)})$
	Elevada	$(X_{(0,0.82)}, 1)$

No primeiro mapa estão representados os dados referentes aos 3 anos, sem discriminação de nenhuma variável ou ajustamento da frequência de sinistros.

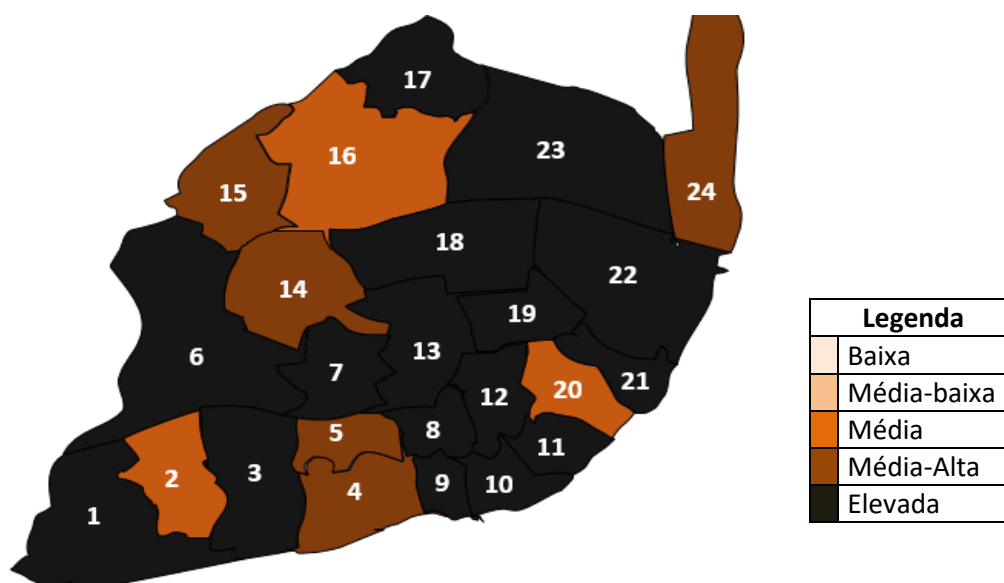


*Mapa 3.1. Portugal Continental (frequência real)*

Podemos verificar que em zonas mais habitacionais existe uma grande concentração de casos com frequência de sinistros mais elevada, como por exemplo a região litoral centro (concelho de Lisboa) e região litoral norte (concelhos do Porto e de Vila Nova de Gaia).

Estas são zonas extremamente importantes porque correspondem a uma grande percentagem de população por área e como tal a uma maior oportunidade de celebração de contratos de seguros. Contudo estas áreas têm elevados custos por sinistros, por isto fui analisar estas regiões ao detalhe para perceber quais são as freguesias mais sensíveis em termos de frequência.

## Concelho de Lisboa

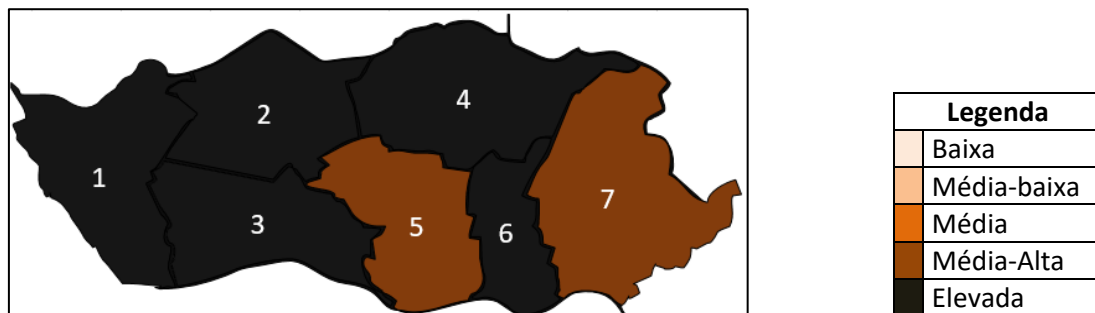


*Mapa 3.2. Concelho de Lisboa repartido por freguesias(frequência real)*

*Tabela 3.3. Correspondência da freguesia (Lisboa)*

Número	Freguesia
1	Belém
2	Ajuda
3	Alcântara
4	Estrela
5	Campo de Ourique
6	Benfica
7	Campolide
8	Santo António
9	Misericórdia
10	Santa Maria Maior
11	São Vicente
12	Arroios
13	Avenidas Novas
14	São Domingos de Benfica
15	Carnide
16	Lumiar
17	Santa Clara
18	Alvalade
19	Areeiro
20	Penha de França
21	Beato
22	Marvila
23	Olivais
24	Parque das Nações

### Concelho do Porto

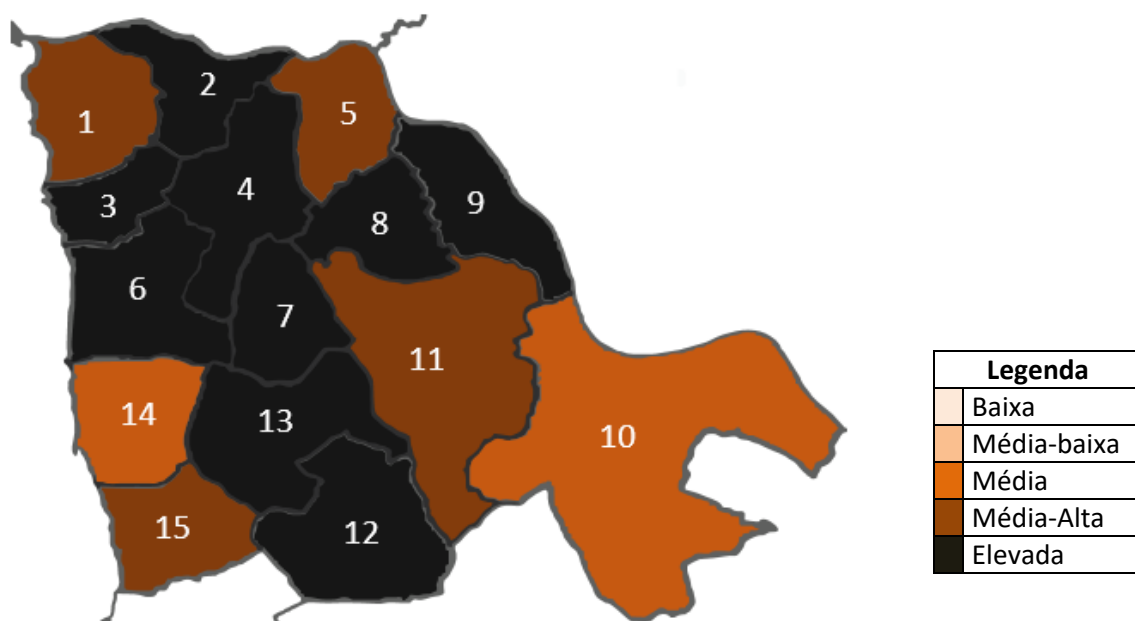


*Mapa 3.3. Concelho do Porto repartido por freguesias (frequência real)*

*Tabela 3.4. Correspondência da freguesia (Porto)*

Número	Freguesia
1	Aldoar, Foz do Douro e Nevogilde
2	Ramalde
3	Lordelo do ouro e Massarelos
4	Maranhos
5	Cedofeita, Santo Ildefonso, Sé, São Nicolau e Vitória
6	Bonfim
7	Campanhã

### Concelho de Vila Nova de Gaia



*Mapa 3.4. Concelho de Vila Nova de Gaia repartido por freguesias (frequência real)*



*Tabela 3.5. Correspondência da freguesia (Vila Nova de Gaia)*

Número	Freguesia
1	Canidelo
2	Santa Marinha e São Pedro da Afurada
3	Madalena
4	Mafamude e Vilar do Paraíso
5	Oliveira do Douro
6	Gulpilhares e Valadares
7	Canelas
8	Vilar de Andorinho
9	Avintes
10	Sandim, Olival, Lever e Crestuma
11	Pedroso e Seixezelo
12	Grijó e Sermonde
13	Serzedo e Perosinho
14	Arcozelo
15	São Félix da Marinha

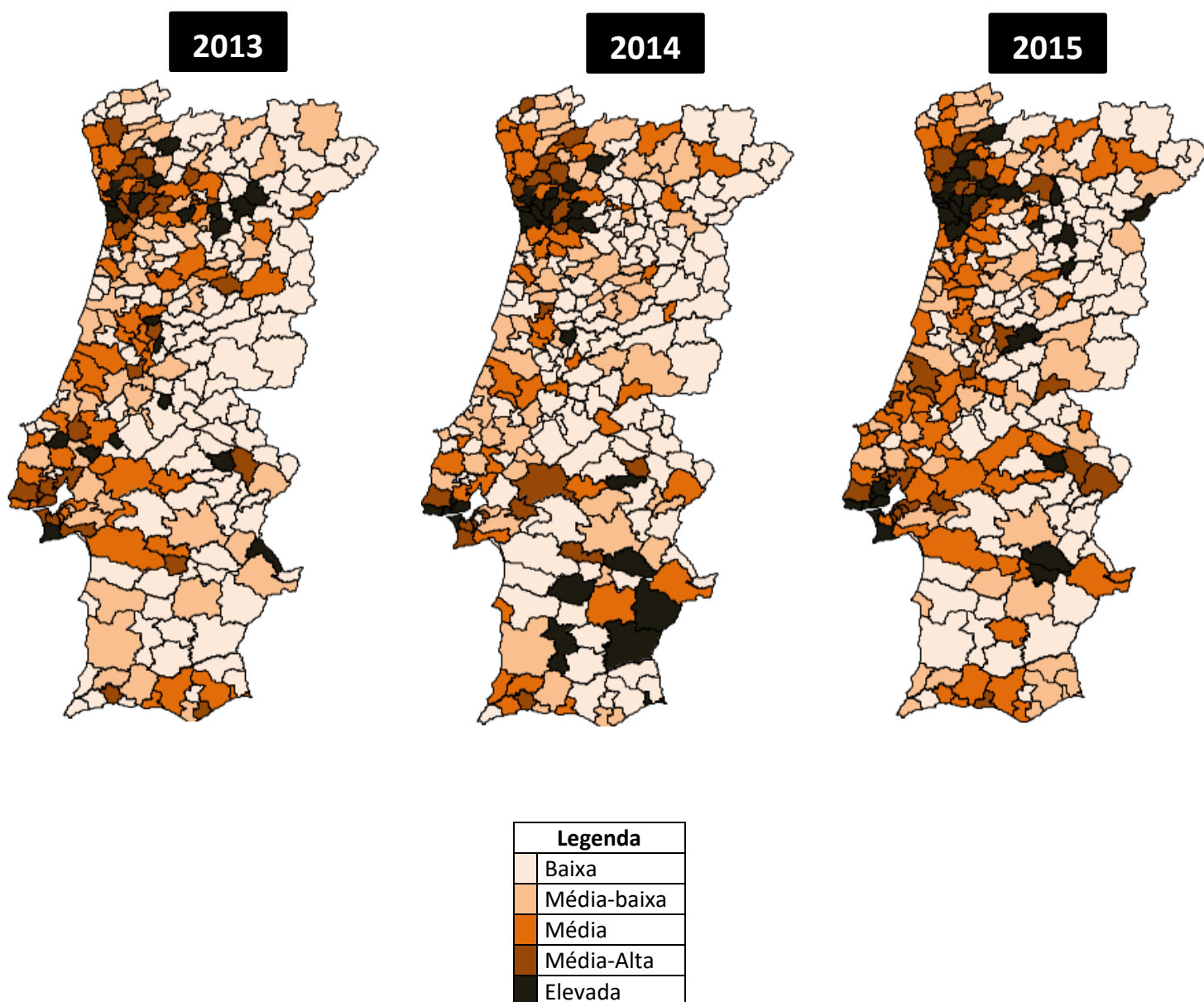
O âmbito deste projeto não é saber quais as causas que provocam maior frequência de sinistros, mas sim saber em que zonas que tal fenómeno ocorre. Após uma breve análise aos mapas pode-se verificar que as freguesias ribeirinhas e as que têm grande número de vias rápidas são as zonas com maior percentagem de sinistros. Isto deve-se ao facto de estas zonas serem muito utilizadas por automobilistas, o que faz com que exista maior probabilidade de um sinistro ocorrer.

### **Evolução anual de Portugal Continental**

Após o mapa global de Portugal Continental ser analisado seria ideal verificar a tendência temporal. Sabe-se que as previsões são uma componente imprescindível na atividade seguradora porque uma seguradora tem a capacidade de identificar e prever quais são os riscos e como tal pode tomar medidas de prevenção para os seus clientes como criar formas de minimizar o prejuízo que alguns riscos e sinistros acarretam para a empresa.

Como tal, é fundamental para uma empresa seguradora, de alguma forma, considerar o comportamento futuro dos elementos de riscos para a base de construção de qualquer tarifa.

A base de dados, como foi dito anteriormente, tem informação temporal dos anos 2013 a 2015 por isso foi verificado qual o comportamento da frequência de sinistros por concelhos ao longo dos anos em Portugal Continental.

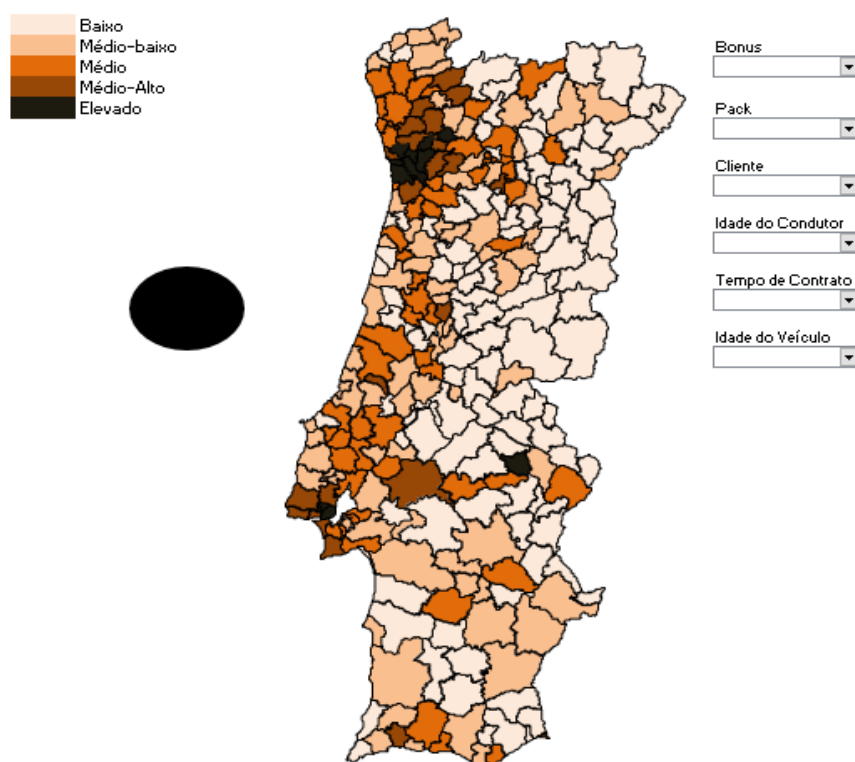


*Mapa 3.5. Evolução anual da frequência real (Portugal Continental - concelhos)*

## Resultados:

- Aumento da frequência nas zonas litorais sendo mais predominante nas regiões de Lisboa e Norte-Litoral;
- Na região de Portalegre também se verifica um aumento ao longo dos anos sendo esta e as regiões referidas no primeiro ponto as mais sensíveis e as que precisam de maior cuidado para a criação das zonas tarifárias porque se preveem ser zonas de alto risco;
- Existem zonas de frequência inconstante (variável) como por exemplo os distritos de Beja e de Bragança, em que não podemos afirmar com todo o rigor como é o comportamento temporal destas regiões.

Para uma melhor compreensão de como a frequência real se comporta em alguns fatores de risco, criei um pequeno simulador em Excel, que devolve a frequência de sinistralidade por concelho conforme as características que queremos observar.



*Mapa 3.6. Simulador de Frequência real de sinistralidade*

Apenas colocou-se 6 características do contrato e do cliente porque a base de dados que suporta o simulador é tão grande que o torna muito pesado e assim ineficaz para uma análise preliminar do comportamento das variáveis em termos de frequência de sinistralidade. Estas 6 variáveis foram as escolhidas porque são as mais fáceis de interpretar e de visualizar para uma pessoa que não conheça as especificidades de um contrato de seguro de automóvel. Pelo que este simulador não só pretende ajudar a área de atuariado como qualquer pessoa que esteja interessada em perceber como a frequência se comporta por concelho.

<b>Bonus</b> Malus Neutro 10%-30% 30%-40% >40%	<b>Pack</b> RC RC+DP SL	<b>Cliente</b> Empresa Particular	<b>Idade do Condutor</b> Jovens Adultos Adultos-Seniores
<b>Tempo de Contrato</b> Danos 1-2anos 3-4anos 5-6anos 7-8anos 9-10anos >10anos		<b>Idade do Veículo</b> 0anos 1-2anos 3-4anos 5-6anos 7-8anos 9-10anos >10anos	

*Figura 3.1. Variáveis do Simulador*

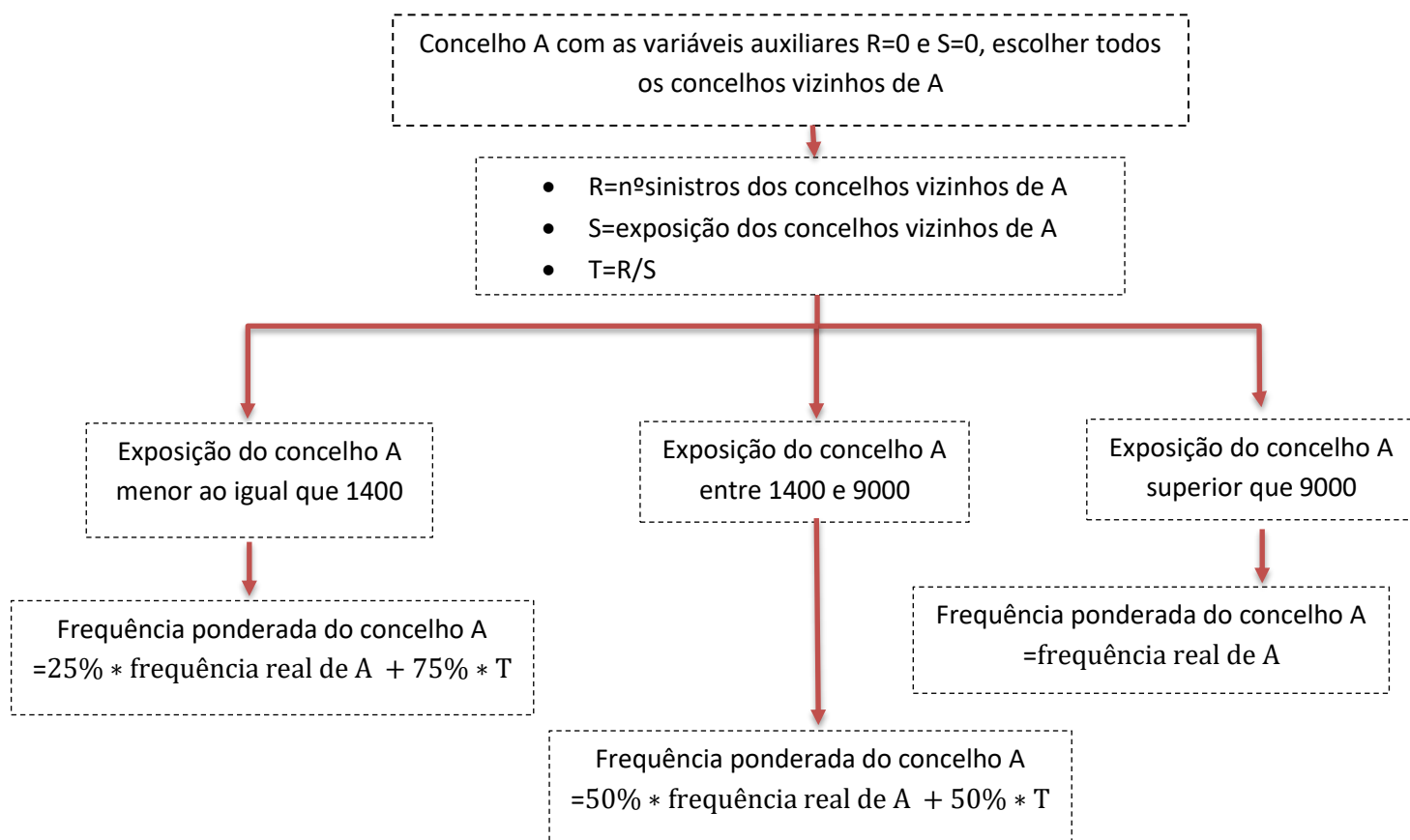
### 3.2.2 Frequência Ponderada

Após a análise da frequência de sinistralidade por concelho o objetivo é criar zonas tarifárias, ou seja, tal como já foi dito anteriormente (página 15) uma zona tarifária corresponde à zona de circulação do veículo, a qual é determinada pelo concelho de residência do segurado. Nesta fase queremos construir 5 zonas tarifárias com base na frequência:

- Baixa
- Média-Baixa
- Média
- Média-Alta
- Elevada

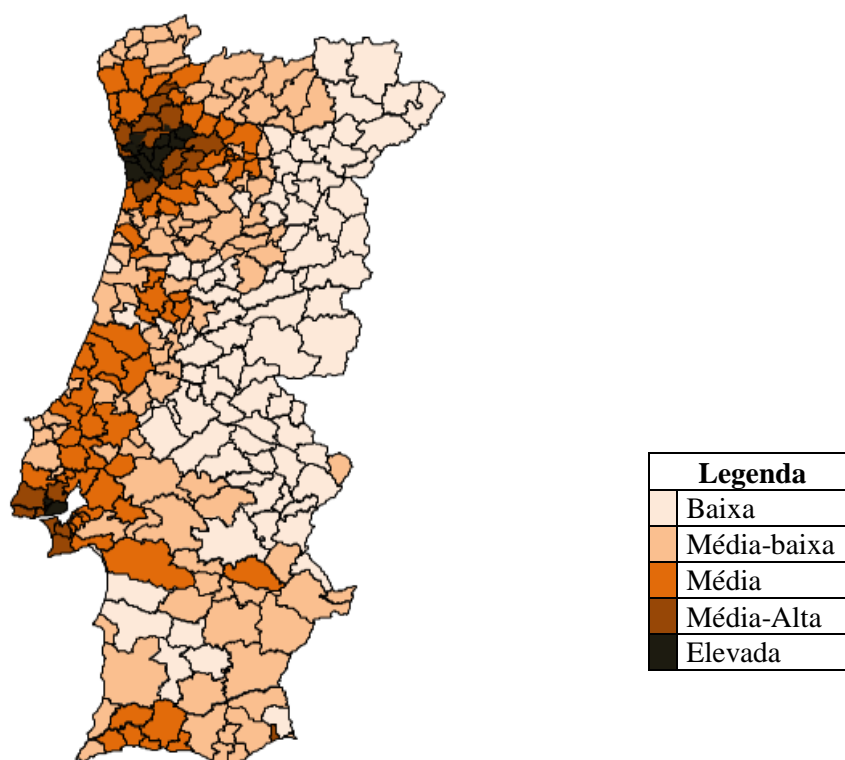
Podia-se utilizar os mapas vistos no subcapítulo acima, no entanto não seria eficaz para uma seguradora ter as zonas tarifárias tão dispersas e confusas. Por isso uma hipótese para a uniformização dos concelhos e a criação da zona tarifária é o cálculo da frequência ponderada.

O esquema abaixo representa como a frequência ponderada de um concelho é calculada:



*Esquema 3.1. Frequência Ponderada*

Após o cálculo das frequências de sinistralidade ponderada de todos os concelhos desenhou-se novamente o mapa de Portugal Continental utilizando a base de dados completa com os 3 anos (2013-2015).



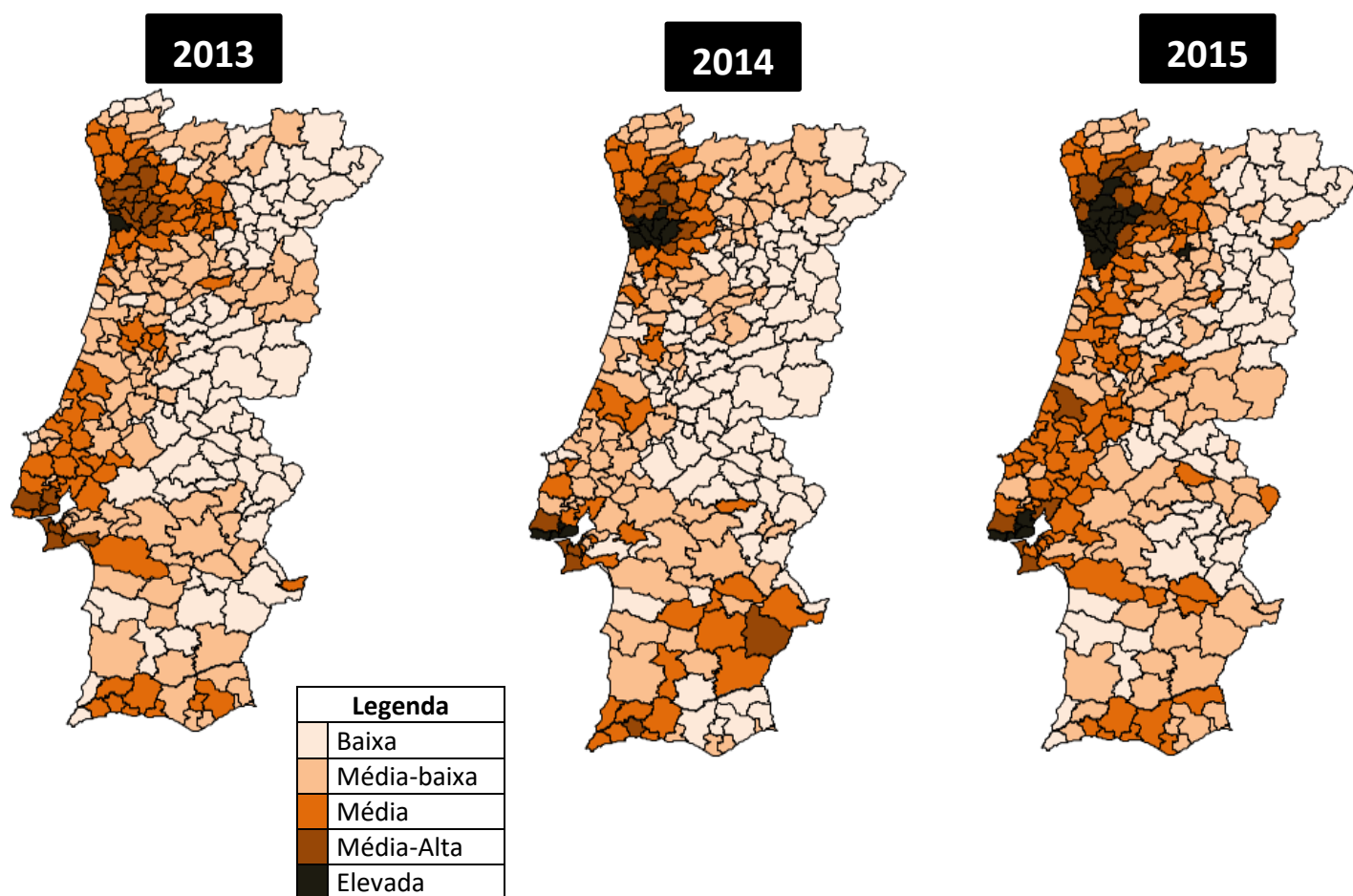
*Mapa 3.7. Portugal Continental (frequência ponderada)*

Pode-se verificar uma clara uniformização das 5 zonas tarifárias. Pelo que:

- Áreas em que existem esporádicas frequências altas com pouca exposição, ou seja, pode-se pôr a hipótese de serem poucos os sinistros mas a frequência alta resulta da pouca exposição que o concelho tem, então estas áreas são mais influenciadas pelos concelhos vizinhos.
- Áreas com exposição elevada nunca são influenciadas pelo concelho vizinho, porque se têm exposição elevada é porque são zonas relevantes para o negócio e como tal precisa-se de ter especial atenção, porque se são zonas com frequência de sinistralidade elevada têm de ser reportadas como tal e a tarifa ser agravada por ser um elevado risco para a seguradora. Por outro lado se tem frequência de sinistralidade baixa não se pode agravar a tarifa porque o concelho vizinho pode ser um caso do referido no ponto anterior.
- Pode-se observar nitidamente que a zona litoral com maior quantidade de zonas de frequência mais elevada é a área com zonas mais escuras enquanto as zonas interiores de Portugal Continental apresentam cores mais claras.

Esta análise oferece uma visão exata de como os concelhos se comportam com eles próprios e com os concelhos ao seu redor.

## Evolução anual de Portugal Continental



*Mapa 3.8. Evolução anual da frequência ponderada (Portugal Continental – concelhos)*

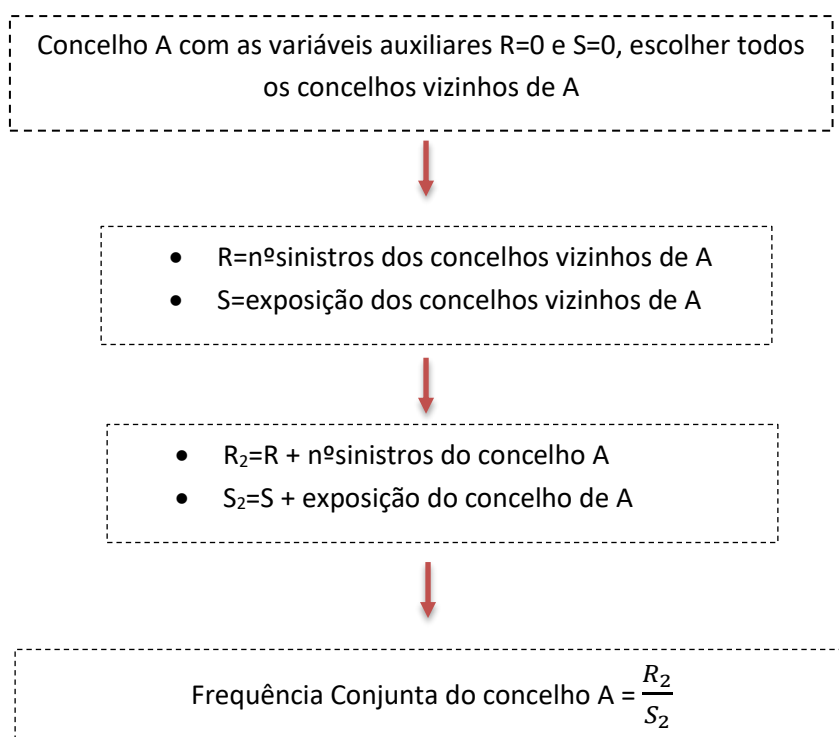
Estes 3 mapas apenas permitem a observação dos concelhos para uma análise temporal porque para a criação de uma tarifa e de zonas tarifárias tem que se ter em avaliação um número considerável de anos e de informação. Se apenas se utilizasse um ano por exemplo 2014 a zona de Beja iria ser agravada quando na realidade a tendência da zona seria uma cor Média-Baixa. O mesmo acontece pelo benefício na zona Litoral-Norte, em que em 2013 a sinistralidade não é muito má. Esta área teria apenas um concelho com frequência de sinistralidade elevada, contudo se observarmos a tendência temporal da região pode se verificar um agravamento substancial da região Litoral-Norte.

No entanto podemos tirar a mesma conclusão do mapa global com a informação completa, zonas com exposição elevada e com sinistralidade elevada são mais influenciadas por si mesmas e por isso encaixam-se num padrão mais elevado, enquanto zonas com exposição baixa são influenciadas pelo comportamento dos seus concelhos vizinhos. Por isso a zona litoral apresenta um agravamento de cor enquanto regiões interiores de Portugal Continental tendem a ter cor mais neutras.

### 3.2.3 Frequência Conjunta

Com a análise realizada no subcapítulo anterior (3.2.2 Frequência Ponderada) consegue-se distinguir visivelmente as 5 zonas tarifárias agrupadas conforme a sua frequência e dos concelhos vizinhos. No entanto esta análise não é 100% eficaz porque se repararmos novamente no mapa 7 existem concelhos isolados na sua própria cor como é o exemplo do concelho de Castro Marim (distrito de Faro) e do concelho Campo Maior (distrito de Portalegre). Para contornar estas falhas do cálculo da frequência Ponderada, calcula-se o que se denominou de frequência conjunta.

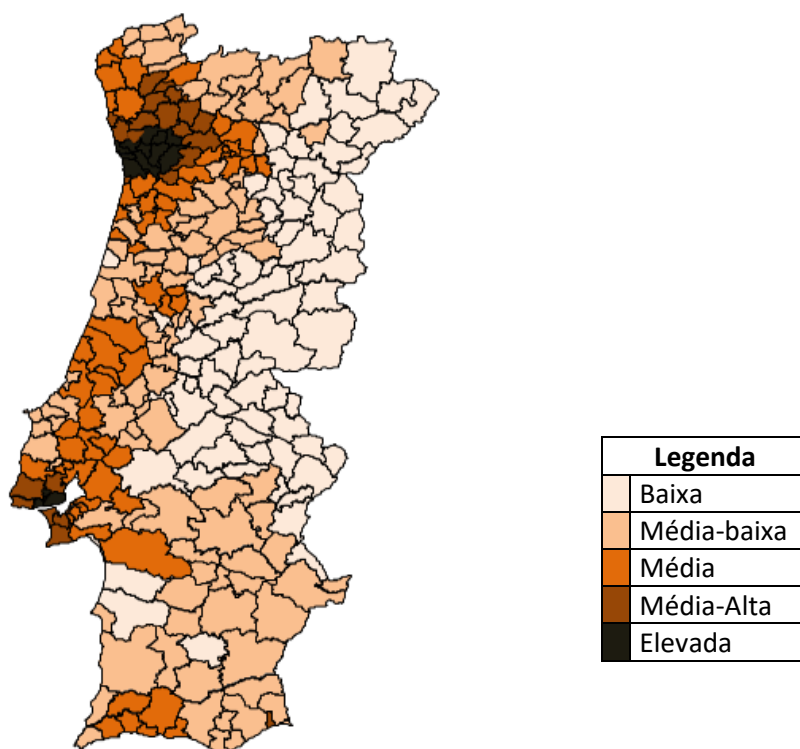
Para o cálculo da frequência conjunta é utilizada toda a informação dos concelhos vizinhos, ou seja, não existe a ponderação percentual com base na exposição do concelho entre a frequência do próprio concelho e a frequência dos concelhos vizinhos. A frequência conjunta de um concelho é a frequência dele e dos seus concelhos vizinhos. Esquematicamente a frequência conjunta é calculada da seguinte forma:



*Esquema 3.2. Frequência Conjunta*



Utilizando os dados completos com os 3 anos (2013-2015) construiu-se novamente o mapa de Portugal Continental agora com a frequência conjunta ilustrada.



*Mapa 3.9. Portugal Continental (frequência conjunta)*

Ao se analisar este mapa pode-se claramente verificar uma uniformização consistente das cores conforme o nível de frequência. No entanto ainda se podem constatar 2 concelhos isolados:

- Ílhavo (distrito de Aveiro): A frequência deste concelho é Média porque tem uma grande relevância do número de sinistros que prevalece sobre o nº de sinistros total (variável  $R_2$ )
- Castro Verde (distrito de Beja): Contrariamente a Ílhavo, Castro Verde tem uma exposição grande para o nº de sinistros ocorridos neste concelho, pelo que do conjunto do concelho de Castro Verde com os concelhos vizinhos irá prevalecer a informação do concelho de Castro Verde.

Com esta análise pode-se verificar o que já se tinha verificado anteriormente, que a zona litoral é a que apresenta cores de frequência mais elevado enquanto a zona interior exhibe cores mais claras.

Em suma esta análise é a forma mais correta e com menor margem de erro de construir zonas tarifárias com base na frequência de sinistros.



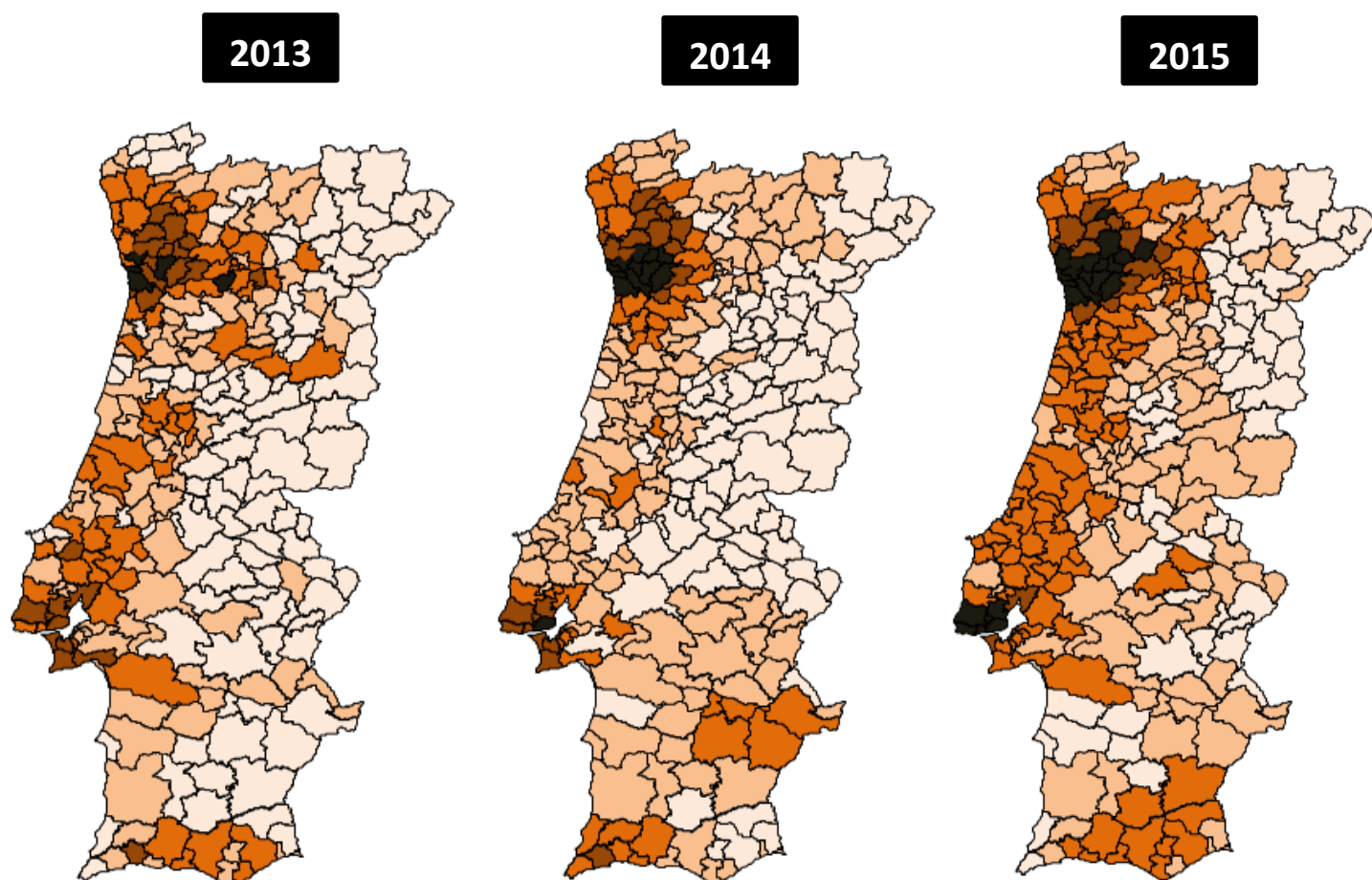
**Tabela 3.6.** Percentagem de concelhos de Portugal Continental por zona tarifária

Frequência de Sinistros	% de concelhos em Portugal Continental
Baixa	27%
Média-Baixa	37%
Média	22%
Média-Alta	9%
Alta	5%

A grande parte dos concelhos concentra-se essencialmente entre o primeiro e o terceiro escalão de frequência (86%). Se aplicássemos estas zonas numa nova criação de tarifa a seguradora estaria a ser mais competitiva nas zonas com menor frequência de sinistros e procuraria gerar mais negócio nestas regiões. Não só o prémio seria mais baixo, como tal seria mais favorável para o cliente, e consequentemente o risco para a empresa seria menor.

Contrariamente nas zonas tarifárias Média-Alta e Alta (14%) a seguradora procuraria aplicar o prémio em sintonia com o risco elevado que iria suportar.

### **Evolução anual de Portugal Continental**

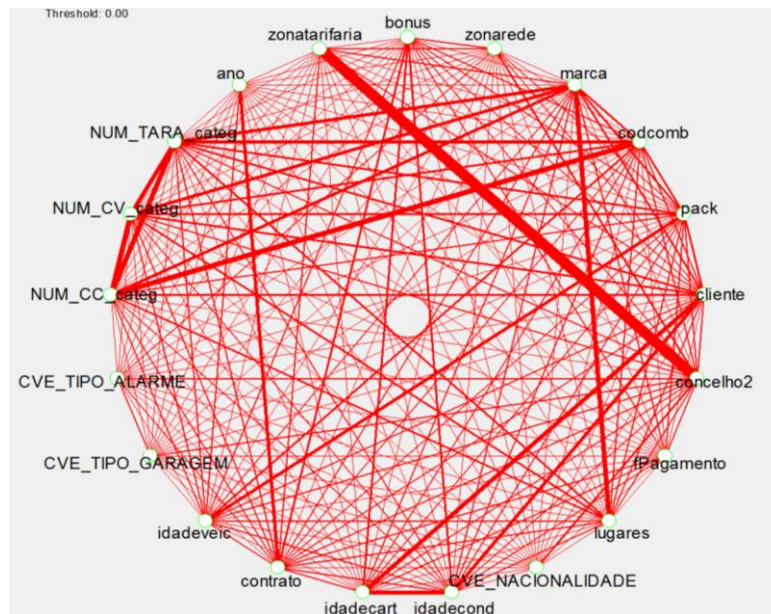


**Mapa 3.10.** Evolução anual da frequência conjunta (Portugal Continental – concelhos)

### 3.3 Modelação

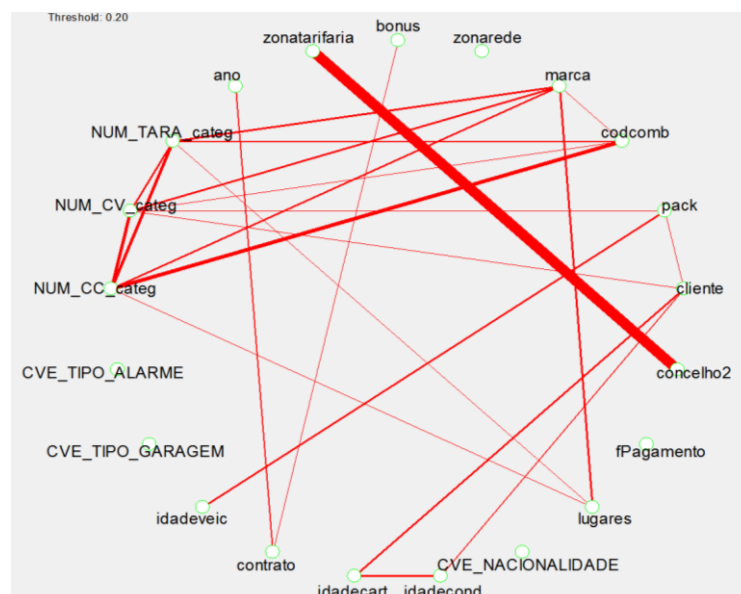
A base de dados criada no *software* SAS foi importada para o *software* Emblem. O Emblem é um *software* de pricing que suporta uma ampla gama de modelos preditivos complexos, permitindo assim ao utilizador aplicar técnicas estatísticas para a modelação.

Em primeiro aplicou-se o método de correlação para testar a multicolinearidade das variáveis.



**Figura 3.2.** Correlações entre as variáveis

Podemos observar que existem correlações elevadas entre variáveis e por isso ir-se-á aplicar um limite de 0.2, e se necessário um limite de correlação de 0.7 para testar a multicolineariedade de variáveis.

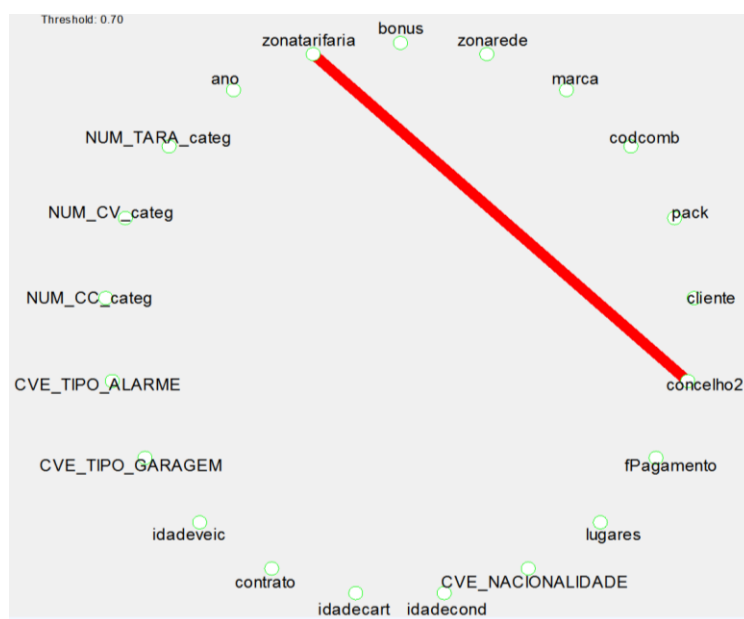


**Figura 3.3.** Correlações entre as variáveis com um limite de correlação igual ou superior a 0.2

Verifica-se que as variáveis mais correlacionadas (linhas vermelhas mais destacadas) são:

- A variável correspondente à zona tarifária (Zonatarifaria) com a variável que determina o concelho (concelho2);
- A variável correspondente ao tipo de combustível (codcomb) com a variável que determina a cilindrada do veículo (NUM\_CC\_categ);
- A variável correspondente à tara do veículo (NUM\_TARA\_categ) com a variável que determina a cilindrada do veículo (NUM\_CC\_categ);
- A variável correspondente à tara do veículo (NUM\_TARA\_categ) com a variável que determina a potência do veículo (NUM\_CV\_categ);
- A variável correspondente à potência do veículo (NUM\_CV\_categ) com a variável que determina a cilindrada do veículo (NUM\_CC\_categ).

Contudo não se pode tirar numa conclusão em concreto e por isso analisa-se a correlação igual ou superior a 0.7.



**Figura 3.4.** Correlações entre as variáveis com um limite de correlação igual ou superior a 0.7

Como a correlação entre a zona tarifária e o concelho é superior ou igual a 0.7, o que faz sentido porque a zona tarifária é criada através do concelho, podemos admitir que existe multicolinearidade entre a variável zonatarifaria e concelho2 por isso a variável zonatarifaria irá ser eliminada do modelo inicial.

O passo seguinte é utilizar os métodos de seleção de variáveis para a construção do modelo final que melhor explica a frequência real de um contrato. Irão serem testados os três métodos e o escolhido será o método que obtenha um menor valor AIC.

### Método Forward

#### **Passo 1:**

Variável	AIC	F
bonus	358755,03	0,00%
zonarede	384441,59	0,00%
marca	384259,13	0,00%
codcomb	384102,36	0,00%
pack	383686,35	0,00%
cliente	384050,38	0,00%
concelho2	382175,59	0,00%
fPagamento	383128,76	0,00%
lugares	384632,22	0,00%
CVE_NACIONALIDADE	384662,27	92,95%
idadecond	383865,99	0,00%
idadecart	383182,52	0,00%
contrato	382304,38	0,00%
idadeveic	384324,32	0,00%
CVE_TIPO_GARAGEM	384588,08	0,00%
CVE_TIPO_ALARME	384660,39	22,08%
NUM_CC_categ	384336,57	0,00%
NUM_CV_categ	384145,13	0,00%
NUM_TARA_categ	384193,71	0,00%
ano	384566,91	0,00%

Como a variável bonus é a variável que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

#### **Passo 2:**

Variável	AIC	F
Current Model Stat	358755,02	0,00%
zonarede	358650,56	0,00%
marca	358506,17	0,00%
codcomb	358515,97	0,00%
pack	358266,32	0,00%
cliente	358514,95	0,00%
concelho2	356765,44	0,00%
fPagamento	358654,27	0,00%
lugares	358730,67	0,00%
CVE_NACIONALIDADE	358753,64	9,48%
idadecond	357867,04	0,00%
idadecart	358483,45	0,00%
contrato	357626,41	0,00%
idadeveic	358727,72	0,00%
CVE_TIPO_GARAGEM	358742,22	0,18%
CVE_TIPO_ALARME	358719,17	0,00%
NUM_CC_categ	358413,32	0,00%
NUM_CV_categ	358465,23	0,00%
NUM_TARA_categ	358463,55	0,00%
ano	358653,85	0,00%

Como a variável marca é a variável que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 3:

Variável	AIC	F
Current Model Stat	358506,17	0,00%
zonarede	358410,22	0,00%
codcomb	358369,73	0,00%
pack	358118,08	0,00%
cliente	358319,40	0,00%
concelho2	356545,69	0,00%
fPagamento	358389,81	0,00%
lugares	358490,92	0,04%
CVE_NACIONALIDADE	358504,21	7,09%
idadecond	357621,72	0,00%
idadecart	358259,53	0,00%
contrato	357361,63	0,00%
idadeveic	358488,21	0,04%
CVE_TIPO_GARAGEM	358493,46	0,19%
CVE_TIPO_ALARME	358474,06	0,00%
NUM_CC_categ	358355,07	0,00%
NUM_CV_categ	358396,64	0,00%
NUM_TARA_categ	358363,05	0,00%
ano	358410,09	0,00%

Como a variável codcomb é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 4:

Variável	AIC	F
Current Model Stat	358369,73	0,00%
zonarede	358271,65	0,00%
pack	358062,47	0,00%
cliente	358213,68	0,00%
concelho2	356380,95	0,00%
fPagamento	358249,62	0,00%
lugares	358366,90	5,98%
CVE_NACIONALIDADE	358367,81	7,22%
idadecond	357406,28	0,00%
idadecart	358117,58	0,00%
contrato	357162,30	0,00%
idadeveic	358372,19	24,73%
CVE_TIPO_GARAGEM	358357,97	0,26%
CVE_TIPO_ALARME	358338,34	0,00%
NUM_CC_categ	358318,89	0,00%
NUM_CV_categ	358314,90	0,00%
NUM_TARA_categ	358307,18	0,00%
ano	358279,90	0,00%

Como a variável pack é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.



### Passo 5:

Variável	AIC	F
Current Model Stat	358062,47	0,00%
zonarede	357980,69	0,00%
cliente	357974,25	0,00%
concelho2	356131,17	0,00%
fPagamento	357980,97	0,00%
lugares	358057,84	2,84%
CVE_NACIONALIDADE	358060,40	6,69%
idadecond	357067,16	0,00%
idadecart	357806,93	0,00%
contrato	356769,87	0,00%
idadeveic	357986,51	0,00%
CVE_TIPO_GARAGEM	358050,93	0,28%
CVE_TIPO_ALARME	358036,73	0,00%
NUM_CC_categ	358013,99	0,00%
NUM_CV_categ	358048,64	0,17%
NUM_TARA_categ	358035,20	0,00%
ano	357970,16	0,00%

Como a variável concelho2 é a variável que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 6:

Variável	AIC	F
Current Model Stat	356131,17	0,00%
zonarede	356094,61	0,00%
cliente	356087,54	0,00%
fPagamento	356071,97	0,00%
lugares	356121,50	0,33%
CVE_NACIONALIDADE	356128,94	6,03%
idadecond	355134,53	0,00%
idadecart	355879,35	0,00%
contrato	354887,73	0,00%
idadeveic	355996,07	0,00%
CVE_TIPO_GARAGEM	356119,74	0,27%
CVE_TIPO_ALARME	356118,75	0,05%
NUM_CC_categ	356080,66	0,00%
NUM_CV_categ	356127,93	4,44%
NUM_TARA_categ	356121,64	0,55%
ano	356050,02	0,00%

Como a variável idadecond é a variável que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 7:**

Variável	AIC	F
Current Model Stat	355134,53	0,00%
zonarede	355101,90	0,00%
cliente	355069,85	0,00%
fPagamento	354970,71	0,00%
lugares	355131,72	5,89%
CVE_NACIONALIDADE	355133,21	9,62%
idadecart	355032,57	0,00%
contrato	354088,11	0,00%
idadeveic	355036,29	0,00%
CVE_TIPO_GARAGEM	355119,43	0,07%
CVE_TIPO_ALARME	355126,13	0,33%
NUM_CC_categ	355101,99	0,00%
NUM_CV_categ	355129,08	2,35%
NUM_TARA_categ	355123,25	0,30%
ano	355045,55	0,00%

Como a variável fpagamento é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 8:**

Variável	AIC	F
Current Model Stat	354970,71	0,00%
zonarede	354937,36	0,00%
cliente	354882,42	0,00%
lugares	354969,26	10,23%
CVE_NACIONALIDADE	354969,24	8,81%
idadecart	354875,27	0,00%
contrato	353919,09	0,00%
idadeveic	354895,13	0,00%
CVE_TIPO_GARAGEM	354956,34	0,09%
CVE_TIPO_ALARME	354963,58	0,57%
NUM_CC_categ	354935,87	0,00%
NUM_CV_categ	354963,29	1,23%
NUM_TARA_categ	354956,26	0,09%
ano	354882,94	0,00%

Como a variável contrato é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 9:**

Variável	AIC	F
Current Model Stat	353919,10	0,00%
zonarede	353887,67	0,00%
cliente	353848,86	0,00%
lugares	353920,44	33,30%
CVE_NACIONALIDADE	353920,29	41,54%
idadecart	353832,43	0,00%
idadeveic	353871,61	0,00%
CVE_TIPO_GARAGEM	353897,31	0,01%
CVE_TIPO_ALARME	353920,99	76,59%
NUM_CC_categ	353901,12	0,05%
NUM_CV_categ	353914,93	3,57%
NUM_TARA_categ	353905,87	0,15%
ano	353841,50	0,00%

Como a variável ano é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 10:**

Variável	AIC	F
Current Model Stat	353841,50	0,00%
zonarede	353810,23	0,00%
cliente	353771,23	0,00%
lugares	353842,79	32,72%
CVE_NACIONALIDADE	353842,73	42,67%
idadecart	353757,07	0,00%
idadeveic	353787,29	0,00%
CVE_TIPO_GARAGEM	353820,52	0,01%
CVE_TIPO_ALARME	353843,41	78,97%
NUM_CC_categ	353822,84	0,04%
NUM_CV_categ	353837,70	4,01%
NUM_TARA_categ	353828,98	0,20%

Como a variável idadecart é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 11:**

Variável	AIC	F
Current Model Stat	353757,07	0,00%
zonarede	353727,18	0,00%
cliente	353718,05	0,00%
lugares	353758,15	29,88%
CVE_NACIONALIDADE	353757,99	34,53%
idadeveic	353703,21	0,00%
CVE_TIPO_GARAGEM	353735,94	0,01%
CVE_TIPO_ALARME	353758,66	55,84%
NUM_CC_categ	353739,45	0,06%
NUM_CV_categ	353753,07	3,75%
NUM_TARA_categ	353744,44	0,19%

Como a variável idadeveic é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.



**Passo 12:**

Variável	AIC	F
Current Model Stat	353703,22	0,00%
zonarede	353672,88	0,00%
cliente	353660,10	0,00%
lugares	353705,98	59,83%
CVE_NACIONALIDADE	353704,49	43,80%
CVE_TIPO_GARAGEM	353681,67	0,01%
CVE_TIPO_ALARME	353705,99	0,00%
NUM_CC_categ	353694,15	0,61%
NUM_CV_categ	353693,73	0,69%
NUM_TARA_categ	353685,37	0,03%

Como a variável CVE\_TIPO\_ALARME é a variável que provoca maior redução na S<sub>Q</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 13:**

Variável	AIC	F
Current Model Stat	353705,17	0,00%
zonarede	353674,80	0,00%
cliente	353662,09	0,00%
lugares	353707,93	59,84%
CVE_NACIONALIDADE	353706,45	43,88%
CVE_TIPO_GARAGEM	353684,22	0,01%
NUM_CC_categ	353696,10	0,61%
NUM_CV_categ	353695,70	0,69%
NUM_TARA_categ	353687,33	0,03%

Como a variável cliente é a variável que provoca maior redução na S<sub>Q</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 14:**

Variável	AIC	F
Current Model Stat	353662,09	0,00%
zonarede	353632,31	0,00%
lugares	353664,62	54,58%
CVE_NACIONALIDADE	353663,44	46,29%
CVE_TIPO_GARAGEM	353640,45	0,01%
NUM_CC_categ	353656,30	1,42%
NUM_CV_categ	353654,83	1,38%
NUM_TARA_categ	353647,31	0,08%

Como a variável zonarede é a variável que provoca maior redução na S<sub>Q</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 15:**

Variável	AIC	F
Current Model Stat	353632,31	0,00%
lugares	353634,82	53,96%
CVE_NACIONALIDADE	353633,70	47,85%
CVE_TIPO_GARAGEM	353610,99	0,01%
NUM_CC_categ	353625,74	1,15%
NUM_CV_categ	353625,01	1,36%
NUM_TARA_categ	353618,35	0,11%

Como a variável CVE\_TIPO\_GARAGEM é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 16:**

Variável	AIC	F
Current Model Stat	353611,09	0,00%
lugares	353612,61	35,81%
CVE_NACIONALIDADE	353611,31	22,37%
NUM_CC_categ	353603,33	0,85%
NUM_CV_categ	353602,74	0,98%
NUM_TARA_categ	353595,32	0,06%

Como a variável NUM\_TARA\_CATEG é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no

modelo.

**Passo 17:**

Variável	AIC	F
Current Model Stat	353595,32	0,00%
lugares	353596,59	43,59%
CVE_NACIONALIDADE	353597,59	0,00%
NUM_CC_categ	353592,87	3,20%
NUM_CV_categ	353596,24	14,54%

A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

Como a variável CVE\_NACIONALIDADE é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada.

**Passo 18:**

Variável	AIC	F
Current Model Stat	353596,68	0,00%
lugares	353599,09	51,67%
NUM_CC_categ	353594,12	3,12%
NUM_CV_categ	353597,60	14,54%

teste F é menor que 5%, então a variável é introduzida no modelo.

Como a variável NUM\_CC\_categ é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do

**Passo 19:**

Variável	AIC	F
Current Model Stat	353592,98	0,00%
lugares	353597,39	72,52%
NUM_CV_categ	353600,53	53,52%

Como as variáveis restantes não provocam uma maior redução na S<sub>Q</sub> então o método acaba com as 18 variáveis e com um AIC=353592.98.

Método *Backward***Passo 1:**

Variável	AIC	F
Current Model Stat	353602,87	0,00%
bonus	377042,46	0,00%
zonarede	353632,47	0,00%
marca	353628,39	0,00%
codcomb	353702,72	0,00%
pack	353908,64	0,00%
cliente	353642,34	0,00%
concelho2	355384,96	0,00%
fPagamento	353761,59	0,00%
lugares	353599,97	46,17%
CVE_NACIONALIDADE	353601,77	38,63%
idadecond	354329,61	0,00%
idadecart	353652,44	0,00%
contrato	354564,85	0,00%
idadeveic	353657,84	0,00%
CVE_TIPO_GARAGEM	353625,03	0,00%
CVE_TIPO_ALARME	353602,20	29,31%
NUM_CC_categ	353600,25	10,02%
NUM_CV_categ	353597,31	42,88%
NUM_TARA_categ	353610,39	1,19%
ano	353683,65	0,00%

O modelo começa com todas as variáveis possíveis.

A variável lugares é a variável com o maior p-value sendo que o valor é superior a 5% então a variável é eliminada do modelo.

## Passo 2:

Variável	AIC	F
Current Model Stat	353599,88	0,00%
bonus	377037,92	0,00%
zonarede	353629,01	0,00%
marca	353632,61	0,00%
codcomb	353701,25	0,00%
pack	353906,06	0,00%
cliente	353639,26	0,00%
concelho2	355381,01	0,00%
fPagamento	353758,73	0,00%
CVE_NACIONALIDADE	353598,42	0,00%
idadecond	354328,24	0,00%
idadecart	353649,41	0,00%
contrato	354563,10	0,00%
idadeveic	353656,88	0,00%
CVE_TIPO_GARAGEM	353621,92	0,00%
CVE_TIPO_ALARME	353599,05	0,00%
NUM_CC_categ	353597,75	10,02%
NUM_CV_categ	353594,14	51,89%
NUM_TARA_categ	353607,48	1,03%
ano	353680,48	0,00%

A variável NUM\_CV\_categ é a variável com o maior p-value sendo que o valor é superior a 5% então a variável é eliminada do modelo.

## Passo 3:

Variável	AIC	F
Current Model Stat	353594,28	0,00%
bonus	377037,42	0,00%
zonarede	353623,48	0,00%
marca	353624,61	0,00%
codcomb	353703,44	0,00%
pack	353903,59	0,00%
cliente	353633,28	0,00%
concelho2	355379,36	0,00%
fPagamento	353753,02	0,00%
CVE_NACIONALIDADE	353592,82	0,00%
idadecond	354322,65	0,00%
idadecart	353643,77	0,00%
contrato	354559,78	0,00%
idadeveic	353654,46	0,00%
CVE_TIPO_GARAGEM	353616,32	0,00%
CVE_TIPO_ALARME	353593,45	0,00%
NUM_CC_categ	353596,83	3,41%
NUM_TARA_categ	353604,74	0,34%
ano	353674,92	0,00%

A variável NUM\_CC\_categ é a variável com o maior p-value sendo que o valor é inferior a 5% então a variável é mantida no modelo.

Como a variável com maior p-value é inferior a 5% então o método termina que se o modelo é composta pelas atuais variáveis com um AIC=353594.28.

## Método Stepwise

### **Passo 1:**

Variável	AIC	F
Current Model Stat	384660,28	0,00%
bonus	358755,03	0,00%
zonarede	384441,59	0,00%
marca	384259,13	0,00%
codcomb	384102,36	0,00%
pack	383686,35	0,00%
cliente	384050,38	0,00%
concelho2	382175,59	0,00%
fPagamento	383128,76	0,00%
lugares	384632,22	0,00%
CVE_NACIONALIDADE	384662,27	92,95%
idadecond	383865,99	0,00%
idadecart	383182,52	0,00%
contrato	382304,38	0,00%
idadeveic	384324,32	0,00%
CVE_TIPO_GARAGEM	384588,08	0,00%
CVE_TIPO_ALARME	384660,39	22,08%
NUM_CC_categ	384336,57	0,00%
NUM_CV_categ	384145,13	0,00%
NUM_TARA_categ	384193,71	0,00%
ano	384566,91	0,00%

Como a variável bonus é a variável que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### **Passo 2:**

Variável	AIC	F
Current Model Stat	358755,02	0,00%
bonus	384660,28	0,00%

Depois de incluída a variável bonus, testar a significância da variável através de um teste F, como a variável é

significativa não é retirada do modelo.

### **Passo 3:**

Variável	AIC	F
Current Model Stat	358755,03	0,00%
zonarede	358650,56	0,00%
marca	358506,17	0,00%
codcomb	358515,97	0,00%
pack	358266,32	0,00%
cliente	358514,95	0,00%
concelho2	356765,44	0,00%
fPagamento	358654,27	0,00%
lugares	358730,67	0,00%
CVE_NACIONALIDADE	358753,64	9,42%
idadecond	357867,04	0,00%
idadecart	358483,45	0,00%
contrato	357626,41	0,00%
idadeveic	358727,72	0,00%
CVE_TIPO_GARAGEM	358742,22	0,18%
CVE_TIPO_ALARME	358719,17	0,00%
NUM_CC_categ	358413,32	0,00%
NUM_CV_categ	358465,23	0,00%
NUM_TARA_categ	358463,55	0,00%
ano	358653,85	0,00%

Como a variável marca é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

#### Passo 4:

Variável	AIC	F
Current Model Stat	358506,17	0,00%
bonus	384258,86	0,00%
marca	358755,01	0,00%

Depois de incluída a variável marca testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são

significativas não irá ser retirada nenhuma variável do modelo.

#### Passo 5:

Variável	AIC	F
Current Model Stat	358506,17	0,00%
zonarede	358410,22	0,00%
codcomb	358369,73	0,00%
pack	358118,08	0,00%
cliente	358319,40	0,00%
concelho2	356545,69	0,00%
fPagamento	358389,81	0,00%
lugares	358490,92	0,04%
CVE_NACIONALIDADE	358504,21	7,09%
idadecond	357621,72	0,00%
idadecart	358259,53	0,00%
contrato	357361,63	0,00%
idadeveic	358488,21	0,04%
CVE_TIPO_GARAGEM	358493,46	0,19%
CVE_TIPO_ALARME	358474,06	0,00%
NUM_CC_categ	358355,07	0,00%
NUM_CV_categ	358396,64	0,00%
NUM_TARA_categ	358363,05	0,00%
ano	358410,09	0,00%

Como a variável codcomb é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

#### Passo 6:

Variável	AIC	F
Current Model Stat	358369,73	0,00%
bonus	383885,63	0,00%
marca	358515,97	0,00%
codcomb	358506,17	0,00%

Depois de incluída a variável codcomb testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada

nenhuma variável do modelo.



### Passo 7:

Variável	AIC	F
Current Model Stat	358369,73	0,00%
zonarede	358271,65	0,00%
pack	358062,47	0,00%
cliente	358213,68	0,00%
concelho2	356380,95	0,00%
fPagamento	358249,62	0,00%
lugares	358366,90	5,98%
CVE_NACIONALIDADE	358367,81	7,21%
idadecond	357406,28	0,00%
idadecart	358117,58	0,00%
contrato	357162,30	0,00%
idadeveic	358372,19	24,72%
CVE_TIPO_GARAGEM	358357,97	0,26%
CVE_TIPO_ALARME	358338,34	0,00%
NUM_CC_categ	358318,89	0,00%
NUM_CV_categ	358314,90	0,00%
NUM_TARA_categ	358307,18	0,00%
ano	358279,90	0,00%

Como a variável pack é a variável, dentro das possíveis, a que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 8:

Variável	AIC	F
Current Model Stat	358062,47	0,00%
bonus	383287,98	0,00%
marca	358165,35	0,00%
codcomb	358118,17	0,00%
pack	358369,73	0,00%

Depois de incluída a variável pack testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do

modelo.

### Passo 9:

Variável	AIC	F
Current Model Stat	358062,47	0,00%
zonarede	357980,69	0,00%
cliente	357974,25	0,00%
concelho2	356131,17	0,00%
fPagamento	357980,97	0,00%
lugares	358057,84	2,84%
CVE_NACIONALIDADE	358060,40	6,69%
idadecond	357067,16	0,00%
idadecart	357806,93	0,00%
contrato	356769,87	0,00%
idadeveic	357986,51	0,00%
CVE_TIPO_GARAGEM	358050,93	0,28%
CVE_TIPO_ALARME	358036,73	0,00%
NUM_CC_categ	358013,99	0,00%
NUM_CV_categ	358048,64	0,17%
NUM_TARA_categ	358035,20	0,00%
ano	357970,16	0,00%

Como a variável concelho2 é a variável, dentro das possíveis, a que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 10:

Variável	AIC	F
Current Model Stat	356131,17	0,00%
bonus	380909,24	0,00%
marca	356228,27	0,00%
codcomb	356214,69	0,00%
pack	356380,27	0,00%
concelho2	358062,47	0,00%

Depois de incluída a variável concelho2 testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

### Passo 11:

Variável	AIC	F
Current Model Stat	356131,17	0,00%
zonarede	356094,61	0,00%
cliente	356087,54	0,00%
fPagamento	356071,97	0,00%
lugares	356121,50	0,33%
CVE_NACIONALIDADE	356128,94	6,02%
idadecond	355134,53	0,00%
idadecart	355879,35	0,00%
contrato	354887,73	0,00%
idadeveic	355996,07	0,00%
CVE_TIPO_GARAGEM	356119,74	0,27%
CVE_TIPO_ALARME	356118,75	0,05%
NUM_CC_categ	356080,66	0,00%
NUM_CV_categ	356127,93	4,44%
NUM_TARA_categ	356121,64	0,55%
ano	356050,02	0,00%

Como a variável idadecond é a variável, dentro das possíveis, a que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 12:

Variável	AIC	F
Current Model Stat	355134,53	0,00%
bonus	380180,20	0,00%
marca	355222,14	0,00%
codcomb	355278,87	0,00%
pack	355423,76	0,00%
concelho2	357067,16	0,00%
idadecond	356131,16	0,00%

Depois de incluída a variável idadecond testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

### Passo 13:

Variável	AIC	F
Current Model Stat	355134,57	0,00%
zonarede	355101,90	0,00%
cliente	355069,85	0,00%
fPagamento	354970,71	0,00%
lugares	355131,72	5,81%
CVE_NACIONALIDADE	355133,21	9,45%
idadecart	355032,57	0,00%
contrato	354088,11	0,00%
idadeveic	355036,29	0,00%
CVE_TIPO_GARAGEM	355119,43	0,07%
CVE_TIPO_ALARME	355126,13	0,32%
NUM_CC_categ	355101,99	0,00%
NUM_CV_categ	355129,08	2,32%
NUM_TARA_categ	355123,25	0,29%
ano	355045,55	0,00%

Como a variável fpagamento é a variável, dentro das possíveis, a que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.



**Passo 14:**

Variável	AIC	F
Current Model Stat	354970,71	0,00%
bonus	379256,48	0,00%
marca	355059,15	0,00%
codcomb	355130,95	0,00%
pack	355210,23	0,00%
concelho2	356865,24	0,00%
fPagamento	355135,19	0,00%
idadecond	356072,06	0,00%

Depois de incluída a variável fpagamento testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

**Passo 15:**

Variável	AIC	F
Current Model Stat	354970,72	0,00%
zonarede	354937,36	0,00%
cliente	354882,42	0,00%
lugares	354969,26	10,19%
CVE_NACIONALIDADE	354969,24	8,77%
idadecart	354875,27	0,00%
contrato	353919,09	0,00%
idadeveic	354895,13	0,00%
CVE_TIPO_GARAGEM	354956,34	0,09%
CVE_TIPO_ALARME	354963,58	0,57%
NUM_CC_categ	354935,87	0,00%
NUM_CV_categ	354963,29	1,23%
NUM_TARA_categ	354956,26	0,09%
ano	354882,94	0,00%

Como a variável contrato é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 16:**

Variável	AIC	F
Current Model Stat	353919,10	0,00%
bonus	378069,51	0,00%
marca	353991,72	0,00%
codcomb	354119,01	0,00%
pack	354225,92	0,00%
concelho2	355765,47	0,00%
fPagamento	354088,78	0,00%
idadecond	354813,06	0,00%
contrato	354970,72	0,00%

Depois de incluída a variável contrato testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

**Passo 17:**

Variável	AIC	F
Current Model Stat	353919,09	0,00%
zonarede	353887,67	0,00%
cliente	353848,86	0,00%
lugares	353920,44	33,34%
CVE_NACIONALIDADE	353920,29	41,61%
idadecart	353832,43	0,00%
idadeveic	353871,61	0,00%
CVE_TIPO_GARAGEM	353897,31	0,01%
CVE_TIPO_ALARME	353920,99	76,84%
NUM_CC_categ	353901,12	0,05%
NUM_CV_categ	353914,93	3,57%
NUM_TARA_categ	353905,87	0,15%
ano	353841,50	0,00%

Como a variável ano é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 18:**

Variável	AIC	F
Current Model Stat	353841,50	0,00%
bonus	378042,19	0,00%
marca	353914,40	0,00%
codcomb	354038,27	0,00%
pack	354150,08	0,00%
concelho2	355693,55	0,00%
fPagamento	354011,26	0,00%
idadecond	354733,20	0,00%
contrato	354882,95	0,00%
ano	353919,10	0,00%

Depois de incluída a variável contrato testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

**Passo 19:**

Variável	AIC	F
Current Model Stat	353841,50	0,00%
zonarede	353810,23	0,00%
cliente	353771,23	0,00%
lugares	353842,79	32,72%
CVE_NACIONALIDADE	353842,73	42,66%
idadecart	353757,07	0,00%
idadeveic	353787,29	0,00%
CVE_TIPO_GARAGEM	353820,52	0,01%
CVE_TIPO_ALARME	353843,41	78,96%
NUM_CC_categ	353822,84	0,04%
NUM_CV_categ	353837,70	4,01%
NUM_TARA_categ	353828,98	0,20%

Como a variável idadecart é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 20:**

Variável	AIC	F
Current Model Stat	353757,07	0,00%
bonus	377339,21	0,00%
marca	353830,64	0,00%
codcomb	353961,95	0,00%
pack	354056,44	0,00%
concelho2	355595,48	0,00%
fPagamento	353920,15	0,00%
idadecond	354546,56	0,00%
idadecart	353841,70	0,00%
contrato	354795,63	0,00%
ano	353832,17	0,00%

Depois de incluída a variável idadecart testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

**Passo 21:**

Variável	AIC	F
Current Model Stat	353756,77	0,00%
zonarede	353727,14	0,00%
cliente	353718,04	0,00%
lugares	353758,15	33,90%
CVE_NACIONALIDADE	353757,99	42,46%
idadeveic	353703,21	0,00%
CVE_TIPO_GARAGEM	353736,11	0,01%
CVE_TIPO_ALARME	353758,95	0,00%
NUM_CC_categ	353738,68	0,05%
NUM_CV_categ	353752,97	3,98%
NUM_TARA_categ	353744,43	0,21%

Como a variável CVE\_TIPO\_ALARME é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 22:**

Variável	AIC	F
Current Model Stat	353758,66	0,00%
bonus	377339,92	0,00%
marca	353832,52	0,00%
codcomb	353963,74	0,00%
pack	354057,39	0,00%
concelho2	355593,98	0,00%
fPagamento	353921,83	0,00%
idadecond	354548,33	0,00%
idadecart	353843,62	0,00%
contrato	354787,68	0,00%
CVE_TIPO_ALARME	353756,81	71,88%
ano	353834,01	0,00%

Depois de incluída a variável CVE\_TIPO\_ALARME testar a significância das restantes variáveis já incluídas através de um teste F. Como a variável CVE\_TIPO\_ALARME não é significativa então é retirada do modelo.

**Passo 23:**

Variável	AIC	F
Current Model Stat	353758,66	0,00%
zonarede	353729,00	0,00%
cliente	353719,98	0,00%
lugares	353760,04	33,94%
CVE_NACIONALIDADE	353759,89	42,65%
idadeveic	353705,40	0,00%
CVE_TIPO_GARAGEM	353737,86	0,01%
NUM_CC_categ	353740,59	0,05%
NUM_CV_categ	353754,87	4,00%
NUM_TARA_categ	353746,33	0,21%

Como a variável idadeveic é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Q</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 24:**

Variável	AIC	F
Current Model Stat	353703,21	0,00%
bonus	377294,24	0,00%
marca	353766,93	0,00%
codcomb	353933,89	0,00%
pack	354056,92	0,00%
concelho2	355566,95	0,00%
fPagamento	353846,01	0,00%
idadecond	354479,11	0,00%
idadecart	353787,50	0,00%
contrato	354705,70	0,00%
idadeveic	353756,81	0,00%
ano	353785,24	0,00%

Depois de incluída a variável idadeveic testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

### Passo 25:

Variável	AIC	F
Current Model Stat	353705,17	0,00%
zonarede	353674,80	0,00%
cliente	353662,09	0,00%
lugares	353707,93	59,84%
CVE_NACIONALIDADE	353706,67	0,00%
CVE_TIPO_GARAGEM	353683,72	0,01%
NUM_CC_categ	353696,03	0,60%
NUM_CV_categ	353695,69	0,69%
NUM_TARA_categ	353687,33	0,03%

Como a variável CVE\_NACIONALIDADE é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 26:

Variável	AIC	F
Current Model Stat	353704,49	0,00%
bonus	377293,35	0,00%
marca	353768,30	0,00%
codcomb	353935,14	0,00%
pack	354058,08	0,00%
concelho2	355567,88	0,00%
fPagamento	353847,34	0,00%
CVE_NACIONALIDADE	353703,31	41,11%
idadecond	354480,48	0,00%
idadecart	353788,77	0,00%
contrato	354704,40	0,00%
idadeveic	353758,05	0,00%
ano	353786,49	0,00%

Depois de incluída a variável CVE\_NACIONALIDADE testar a significância das restantes variáveis já incluídas através de um teste F. Como a variável CVE\_NACIONALIDADE não é significativa então é retirada do modelo.

### Passo 27:

Variável	AIC	F
Current Model Stat	353706,45	0,00%
zonarede	353676,13	0,00%
cliente	353663,44	0,00%
lugares	353709,21	59,96%
CVE_TIPO_GARAGEM	353684,88	0,00%
NUM_CC_categ	353697,57	0,64%
NUM_CV_categ	353697,00	0,70%
NUM_TARA_categ	353688,58	0,03%

Como a variável cliente é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 28:

Variável	AIC	F
Current Model Stat	353660,10	0,00%
bonus	377122,01	0,00%
marca	353722,98	0,00%
codcomb	353879,06	0,00%
pack	353982,71	0,00%
cliente	353703,22	0,00%
concelho2	355492,20	0,00%
fPagamento	353816,88	0,00%
idadecond	354385,04	0,00%
idadecart	353710,69	0,00%
contrato	354641,03	0,00%
idadeveic	353718,10	0,00%
ano	353742,53	0,00%

Depois de incluída a variável cliente testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

### Passo 29:

Variável	AIC	F
Current Model Stat	353662,09	0,00%
zonarede	353632,31	0,00%
lugares	353664,62	54,58%
CVE_TIPO_GARAGEM	353640,63	0,00%
NUM_CC_categ	353655,79	1,24%
NUM_CV_categ	353654,76	1,35%
NUM_TARA_categ	353647,31	0,08%

Como a variável zonarede é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

### Passo 30:

Variável	AIC	F
Current Model Stat	353630,34	0,00%
bonus	377107,60	0,00%
zonarede	353660,14	0,00%
marca	353692,88	0,00%
codcomb	353849,56	0,00%
pack	353956,38	0,00%
cliente	353672,87	0,00%
concelho2	355432,33	0,00%
fPagamento	353788,01	0,00%
idadecond	354352,06	0,00%
idadecart	353679,56	0,00%
contrato	354608,70	0,00%
idadeveic	353689,10	0,00%
ano	353712,70	0,00%

Depois de incluída a variável zonarede testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

### Passo 31:

Variável	AIC	F
Current Model Stat	353630,35	0,00%
lugares	353632,86	54,11%
CVE_TIPO_GARAGEM	353609,22	0,01%
NUM_CC_categ	353623,99	1,31%
NUM_CV_categ	353623,34	1,49%
NUM_TARA_categ	353615,62	0,08%

Como a variável CVE\_TIPO\_GARAGEM é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é introduzida

no modelo.

### Passo 32:

Variável	AIC	F
Current Model Stat	353609,22	0,00%
bonus	377081,38	0,00%
zonarede	353637,71	0,00%
marca	353670,74	0,00%
codcomb	353826,60	0,00%
pack	353934,94	0,00%
cliente	353650,54	0,00%
concelho2	355410,14	0,00%
fPagamento	353764,65	0,00%
idadecond	354332,29	0,00%
idadecart	353657,32	0,00%
contrato	354593,91	0,00%
idadeveic	353667,59	0,00%
CVE_TIPO_GARAGEM	353630,35	0,01%
ano	353690,72	0,00%

Depois de incluída a variável CVE\_TIPO\_GARAGEM testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

**Passo 33:**

Variável	AIC	F
Current Model Stat	353608,28	0,00%
lugares	353610,71	52,19%
NUM_CC_categ	353601,76	1,25%
NUM_CV_categ	353601,25	1,48%
NUM_TARA_categ	353592,45	0,04%

introduzida no modelo.

Como a variável NUM\_TARA\_categ é a variável, dentro das possíveis, a que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F é menor que 5%, então a variável é

**Passo 34:**

Variável	AIC	F
Current Model Stat	353593,66	0,00%
bonus	377059,07	0,00%
zonarede	353623,08	0,00%
marca	353640,34	0,00%
codcomb	353752,26	0,00%
pack	353910,86	0,00%
cliente	353632,96	0,00%
concelho2	355383,51	0,00%
fPagamento	353751,50	0,00%
idadecond	354322,35	0,00%
idadecart	353643,53	0,00%
contrato	354575,65	0,00%
idadeveic	353658,32	0,00%
CVE_TIPO_GARAGEM	353615,58	0,01%
NUM_TARA_categ	353609,21	0,06%
ano	353674,55	0,00%

Depois de incluída a variável NUM\_TARA\_categ testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

**Passo 35:**

Variável	AIC	F
Current Model Stat	353596,68	0,00%
lugares	353599,09	51,66%
NUM_CC_categ	353594,29	3,54%
NUM_CV_categ	353597,89	15,67%

é menor que 5%, então a variável é introduzida no modelo.

Como a variável NUM\_CC\_categ é a variável, dentro das possíveis, a que provoca maior redução na SQe é essa a variável a ser testada. A redução é significativa porque o p-value do teste F



**Passo 36:**

Variável	AIC	F
Current Model Stat	353593,02	0,00%
bonus	377038,29	0,00%
zonarede	353623,48	0,00%
marca	353624,71	0,00%
codcomb	353703,73	0,00%
pack	353903,83	0,00%
cliente	353633,04	0,00%
concelho2	355379,21	0,00%
fPagamento	353753,01	0,00%
CVE_NACIONALIDADE	353592,82	0,00%
idadecond	354322,65	0,00%
idadecart	353643,77	0,00%
contrato	354560,13	0,00%
idadeveic	353654,51	0,00%
CVE_TIPO_GARAGEM	353616,32	0,00%
CVE_TIPO_ALARME	353593,45	0,00%
NUM_CC_categ	353596,83	2,51%
NUM_TARA_categ	353604,74	0,21%
ano	353674,92	0,00%

Depois de incluída a variável NUM\_CC\_categ testar a significância das restantes variáveis já incluídas através de um teste F. Como todas as variáveis são significativas não irá ser retirada nenhuma variável do modelo.

**Passo 37:**

Variável	AIC	F
Current Model Stat	353594,11	0,00%
lugares	353597,44	0,00%
NUM_CV_categ	353600,01	45,82%

Como a variável lugares é a variável, dentro das possíveis, a que provoca maior redução na S<sub>Qe</sub> é essa a variável a ser testada. A redução é significativa porque

o p-value do teste F é menor que 5%, então a variável é introduzida no modelo.

**Passo 38:**

Variável	AIC	F
Current Model Stat	353596,18	0,00%
bonus	377041,81	0,00%
zonarede	353626,67	0,00%
marca	353620,61	0,00%
codcomb	353704,50	0,00%
pack	353906,67	0,00%
cliente	353636,43	0,00%
concelho2	355382,98	0,00%
fPagamento	353755,84	0,00%
lugares	353594,21	87,65%
CVE_NACIONALIDADE	353595,90	0,00%
idadecond	354324,00	0,00%
idadecart	353647,12	0,00%
contrato	354561,48	0,00%
idadeveic	353656,24	0,00%
CVE_TIPO_GARAGEM	353619,45	0,00%
CVE_TIPO_ALARME	353596,61	0,00%
NUM_CC_categ	353599,24	3,02%
NUM_TARA_categ	353607,74	0,22%
ano	353678,10	0,00%

Depois de incluída a variável lugares testar a significância das restantes variáveis já incluídas através de um teste F. Como a variável lugares não é significativa então é retirada do modelo.

**Passo 39:**

Variável	AIC	F
Current Model Stat	353597,28	0,00%
NUM_CV_categ	353603,03	52,04%

Como a variável NUM\_CV\_categ não provoca maior redução na S<sub>Qe</sub> porque o p-value do teste F é maior

que 5%, então a variável não é introduzida no modelo. Como já não existem mais variáveis possíveis para introduzir, o método termina sendo que o modelo final tem as seguintes variáveis:

Variável	AIC	F
Current Model Stat	353590,98	0,00%
bonus	377038,36	0,00%
zonarede	353620,35	0,00%
marca	353621,43	0,00%
codcomb	353700,45	0,00%
pack	353901,31	0,00%
cliente	353630,19	0,00%
concelho2	355378,18	0,00%
fPagamento	353750,17	0,00%
idadecond	354319,72	0,00%
idadecart	353640,60	0,00%
contrato	354568,49	0,00%
idadeveic	353651,49	0,00%
CVE_TIPO_GARAGEM	353612,98	0,01%
NUM_CC_categ	353594,62	2,41%
NUM_TARA_categ	353601,73	0,37%
ano	353671,85	0,00%

O modelo final tem 16 variáveis com um AIC=353590.98.

Após a aplicação dos três métodos conclui-se:

*Tabela 3.7. Escolha dos métodos*

Método	Nº de variáveis final	AIC final
Forward	18	353592.98
Backwards	18	353594.28
Stepwise	16	353590.98

O modelo final escolhido é o modelo determinado pelo método de stepwise porque é o que apresenta um AIC menor, e assim as variáveis escolhidas por este método são as que melhor explicam a frequência de sinistros.



### 3.4 Modelo Final

Do modelo final definido pelo método de *stepwise* resultaram as seguintes estimativas do parâmetro  $\beta$ :

- $\beta_0 = -2.3401$

**Tabela 3.8.** Estimativa dos coeficiente: da variável bonus

bonus	
Variável	$\beta$
Malus	2,3401
Neutro	0,8340
10%	0,6343
20%	1,0088
25%	1,1501
30%	-0,0649
32,5%	-0,2340
35%	0,5615
37,5%	-0,4671
40%	0,0000
42,5%	-0,5652
45%	-0,6712
47,5%	-0,7930
50%	-0,9608
+50%	-1,1868

**Tabela 3.9.** Estimativa dos coeficientes da variável zonarede

zonarede	
Variável	$\beta$
B2B	0,3459
Corretores	0,0540
Multiaccess	-0,6036
Private	-0,0295
Residual	0,2084
Território	0,0000
outros	-0,0726

**Tabela 3.10.** Estimativa dos coeficientes da variável codcomb

codcomb	
Variável	$\beta$
D	0,0000
E	0,4042
G	-0,1435
M	0,7377

**Tabela 3.11.** Estimativa dos coeficientes da variável idadecond

idadecond	
Variável	$\beta$
01-[0,25[	-0,1915
02-[25,35[	-0,1724
03-[35,45[	0,0000
04-[45,55[	0,1418
05-[55,65[	0,1052
06-[65,75[	0,1622
07-[75,+oo	0,4914
99-SEMINFO	-1,7586

**Tabela 3.12.** Estimativa dos coeficientes da variável idadecart

idadecart	
Variável	$\beta$
<1ano	0,6746
1 ano	0,5115
2 anos	0,2531
3 a 5 anos	0,1011
6 a 10 anos	0,0515
11 a 15 anos	0,0536
+15 anos	0,0000
99-SEMINFO	-13,9838

**Tabela 3.13.** Estimativa dos coeficientes da variável contrato

contrato	
Variável	$\beta$
<1ano	0,0994
1 ano	0,0295
2 anos	-0,0567
3 a 5 anos	0,0000
6 a 10 anos	0,0030
11 a 15 anos	0,4083
16 a 29 anos	0,4902
30 a 50 anos	0,5185
+50 anos	0,6386

**Tabela 3.14.** Estimativa dos coeficientes da variável fpagamento

fpagamento	
Variável	$\beta$
Anual	0,0000
Mensal	0,1271
Semestral	0,1159
Trimestral	0,1616

**Tabela 3.15.** Estimativa dos coeficientes da variável cliente

cliente	
Variável	$\beta$
Empresa	0,1351
Particular	0,0000

**Tabela 3.16.** Estimativa dos coeficientes da variável pack

pack	
Variável	$\beta$
RC	0,0000
RC+DP	0,2220
RC+IRE+FR	-0,0173
SL	-0,3067

**Tabela 3.17.** Estimativa dos coeficientes da variável marca (1)

marca	
Variável	$\beta$
ALFA ROMEO	0,0604
AUDI	0,0187
BEDFORD	0,5912
BMW	0,1050
CHEVROLET	0,0849
CHRYSLER	0,0644
CITROEN	0,0165
DACIA	-0,2110
DAEWOO	0,0052
DAIHATSU	-0,0211
DATSUN	-0,8671
DODGE	-0,1922
FERRARI	-0,1747
FIAT	-0,0214
FORD	-0,0070
HONDA	-0,0409
HYUNDAI	-0,0623
ISUZU	-0,8702
JAGUAR	-0,2004
JEEP	0,1681

**Tabela 3.18.** Estimativa dos coeficientes da variável marca (2)

marca	
Variável	$\beta$
KIA	-0,0615
LAND ROVER	0,0987
LEXUS	-0,1313
MAZDA	-0,0433
MERCEDES-BENZ	-0,0258
MG	-0,1361
MICRO COMPACT C	0,0577
MINI	-0,1536
MITSUBISHI	0,0919
NISSAN	0,0635
OPEL	-0,0146
OUTRA	-0,3847
PEUGEOT	0,0088
PORSCHE	-0,0835
RANGE ROVER	-0,4957
RENAULT	0,0000
ROVER	-0,0668
SAAB	-0,2124
SEAT	0,0327
SKODA	0,0536
SMART	-0,1169

**Tabela 3.19.** Estimativa dos coeficientes da variável marca (3)

marca	
Variável	$\beta$
SSANGYONG	-0,2091
SUZUKI	-0,0935
TOYOTA	0,0046
UMM	-0,9118
VOLKSWAGEN	0,0163
VOLVO	0,0618

**Tabela 3.20.** Estimativa dos coeficientes da variável concelho2 (1)

concelho2	
Variável	$\beta$
ABRANTES	-0,8462
AGUIAR DA BEIRA	-0,8935
ALANDROAL	-0,7029
ALBERGARIA -A - VELHA	-0,4118
ALBUFEIRA	-0,4026
ALCACÉR DO SAL	-0,4248
ALCANENA	-0,6190
ALCOBAÇA	-0,2603
ALCOCHETE	-0,2107
ALENQUER	-0,3612
ALIJO	-0,9267
ALJEZUR	-0,5419
ALJUSTREL	-0,6864
ALMADA	-0,1087
ALMEIDA	-0,7571
ALMEIRIM	-0,4468
ALMODÔVAR	-1,1205
ALPIARÇA	-0,7396
ALTER DO CHÃO	-0,7860
ALVAIÀZERE	-0,1225

**Tabela 3.21.** Estimativa dos coeficientes da variável concelho2 (2)

concelho2	
Variável	$\beta$
BELMONTE	-0,7872
BENAVENTE	-0,4225
BOMBARRAL	-0,3017
BORBA	-0,6146
BOTICAS	-0,6166
BRAGA	-0,1171
BRAGANÇA	-0,5995
CABECEIRAS DE BASTO	-0,5080
CADAVAL	-0,4927
CALDAS DA RAINHA	-0,3245
CAMINHA	-0,6077
CAMPO MAIOR	-0,7847
CANTANHEDE	-0,4201
CARRAZEDA DE ANSIÃES	-0,5126
CARREGAL DO SAL	-0,6126
CARTAXO	-0,3951
CASCAIS	-0,2034
CASTANHEIRA DE PÊRA	-0,5235
CASTELO BRANCO	-0,3883
CASTELO DE PAIVA	-0,0478

**Tabela 3.22.** Estimativa dos coeficientes da variável concelho2 (3)

concelho2	
Variável	$\beta$
CASTELO DE VIDE	-0,7109
CASTRO DAIRE	-0,5441
CASTRO VERDE	-0,5830
CELORICO DA BEIRA	-21,2165
CELORICO DE BASTO	-0,4856
CHAMUSCA	-0,7836
CHAVES	-0,3781
CINFÃES	-0,2512
COIMBRA	-0,2559
CONDEIXA -A - NOVA	-0,2613
CORUCHE	-0,2696
COVILHÃ	-0,4516
CRATO	-1,2111
ELVAS	-0,2176
ENTRONCAMENTO	-0,4686
ESPINHO	-0,1492
ESPOSENDE	-0,2982
ESTARREJA	-0,5810
FAFE	-0,4862
FARO	-0,4559

**Tabela 3.63.** Estimativa dos coeficientes da variável concelho2 (4)

concelho2	
Variável	$\beta$
ALVITO	-0,8316
AMADORA	-0,0873
AMARANTE	-0,0720
AMARES	-0,2992
ANADIA	-0,3483
ANSIÃO	-0,5115
ARCOS DE VALDEVEZ	-0,6316
ARGANIL	-0,7161
AROUCA	-0,2466
ARRAIOS	-0,7309
ARRONCHES	-0,6391
ARRUDA DOS VINHOS	-0,5186
AVEIRO	-0,2605
AVIS	-1,8053
AZAMBUJA	-0,3611
BAIÃO	-0,3922
BARCELOS	-0,1803
BARREIRO	-0,2558
BATALHA	0,0489
BEJA	-0,3876

**Tabela 3.24.** Estimativa dos coeficientes da variável concelho2 (5)

concelho2	
Variável	$\beta$
LISBOA	-0,1432
LOULÉ	-0,4598
LOURES	-0,1650
LOURINHÃ	-0,4018
LOUSADA	0,0903
LOUSÃ	-0,0520
MACEDO DE CAVALEIROS	-0,4178
MAFRA	-0,5144
MAIA	0,0246
MANGUALDE	-0,2178
MANTEIGAS	-0,3641
MARCO DE CANAVESES	-0,1185
MARINHA GRANDE	-0,2716
MARVÃO	-0,3978
MATOSINHOS	-0,0663
MEALHADA	-0,3438
MELGAÇO	-0,3540
MESÃO FRIO	-0,4574
MIRA	-0,5616
MIRANDA DO CORVO	-0,2380

**Tabela 3.25.** Estimativa dos coeficientes da variável concelho2 (6)

concelho2	
Variável	$\beta$
MIRANDA DO DOURO	-0,8729
MIRANDELA	-0,4236
MOGADOURO	-0,7212
MOITA	-0,3745
MONCHIQUE	-0,1233
MONDIM DE BASTO	-0,5883
MONFORTE	-0,2053
MONTALEGRE	-0,8506
MONTE MOR - O - NOVO	-0,6055
MONTE MOR - O - VELHO	-0,5097
MONTIJO	-0,3448
MONÇÃO	-0,4908
MORA	-0,3382
MORTÁGUA	-0,8680
MOURA	-0,4739
MURTOSA	-0,7724
MURÇA	-0,6361
NAZARÉ	-0,4182
NELAS	-0,4496
NISA	-0,8249

**Tabela 3.26.** Estimativa dos coeficientes da variável concelho2 (7)

concelho2	
Variável	$\beta$
FELGUEIRAS	0,0630
FERREIRA DO ALENTEJO	-0,3916
FERREIRA DO ZÉZERE	-0,2627
FIGUEIRA DA FOZ	-0,3147
FIGUEIRÓ DOS VINHOS	-0,6010
FORNOS DE ALGODRES	-0,3031
FREIXO DE ESPADA À CINTA	-0,4853
FRONTEIRA	0,2270
FUNDÃO	-0,4737
GOLEGÃ	-0,7225
GONDOMAR	0,1135
GOUVEIA	-0,4088
GRÂNDOLA	-0,6129
GUARDA	-0,4097
GUIMARÃES	-0,1735
IDANHA - A - NOVA	-0,7257
LAGOA	-0,4790
LAGOS	-0,4813
LAMEGO	-0,2913
LEIRIA	-0,1245

**Tabela 3.27.** Estimativa dos coeficientes da variável concelho2 (8)

concelho2	
Variável	$\beta$
ODEMIRA	-0,3760
ODIVELAS	-0,1032
OEIRAS	-0,1985
OLEIROS	-0,5428
OLHÃO	-0,4002
OLIVEIRA DE AZEMÉIS	-0,2073
OLIVEIRA DE FRADES	-0,4601
OLIVEIRA DO BAIRRO	-0,3056
OLIVEIRA DO HOSPITAL	-0,5269
OURIQUE	-0,7632
OURÉM	-0,3790
OVAR	-0,3351
PALMELA	-0,5716
PAREDES	0,1005
PAÇOS DE FERREIRA	0,0972
PEDROGÃO GRANDE	-0,6098
PENACOVA	-0,5019
PENAFIEL	0,0080
PENALVA DO CASTELO	-0,4179
PENAMACOR	-1,1321

**Tabela 3.28.** Estimativa dos coeficientes da variável concelho2 (9)

concelho2	
Variável	$\beta$
PENELA	-0,5938
PENICHE	-0,4838
PESO DA RÉGUA	-0,3916
PINHEL	-0,7204
POMBAL	-0,2746
PONTE DA BARCA	-0,5691
PONTE DE LIMA	-0,2308
PONTE DE SOR	-0,6370
PORTALEGRE	-0,5572
PORTEL	-0,0062
PORTIMÃO	-0,2785
PORTO	0,0528
PORTO DE MÓS	-0,3102
PÓVOA DE LANHOSO	-0,3957
PÓVOA DE VARZIM	-0,2296
RESENDE	-0,1695
RIBEIRA DE PENA	-0,6628
RIO MAIOR	-0,4281
SABROSA	-0,6224
SABUGAL	-0,7524

**Tabela 3.29.** Estimativa dos coeficientes da variável concelho2 (10)

concelho2	
Variável	$\beta$
SALVATERRA DE MAGOS	-0,3904
SANTA COMBA DÃO	-0,5812
SANTA MARIA DA FEIRA	-0,0187
SANTA MARTA DE PENAGUIÃO	-0,8614
SANTARÉM	-0,2704
SANTIAGO DO CACÉM	-0,5119
SANTO TIRSO	-0,1631
SEIA	-0,4372
SEIXAL	-0,2256
SERNANCELHE	-0,2807
SERTÃO	-0,7039
SESIMBRA	-0,1060
SETÚBAL	-0,2157
SEVER DO VOUGA	-0,5998
SILVES	-0,4541
SINES	-0,4851
SINTRA	-0,1915
SOBRAL DE MONTE AGRAÇO	-0,5571
SOURE	-0,5595
SOUSEL	-0,0869

**Tabela 3.30.** Estimativa dos coeficientes da variável concelho2 (11)

concelho2	
Variável	$\beta$
VILA NOVA DE PAIVA	-0,9340
SARDOAL	-0,7973
VIDIGUEIRA	-0,5865
ESTREMOZ	-0,7656
VIEIRA DO MINHO	-0,1763
PROENÇA - A - NOVA	-0,7973
GÓIS	-0,6657
VILA DO BISPO	-0,6994
VILA DO CONDE	-0,1844
VILA FRANCA DE XIRA	-0,2097
VILA NOVA DA BARQUINHA	-0,8259
PAREDES DE COURA	-0,9106
VILA NOVA DE GAIA	0,0000
ARMAMAR	-0,4901
VILA NOVA DE POIARES	-0,4068
VILA NOVA FAMALICÃO	-0,0776
VILA POUCA DE AGUIAR	-0,7515
VILA REAL	-0,3492
VILA REAL DE SANTO ANTÓNIO	-0,1968
VILA DE REI	-0,7973

**Tabela 3.31.** Estimativa dos coeficientes da variável concelho2 (12)

concelho2	
Variável	$\beta$
SÁTÃO	-0,4752
SÃO BRÁS DE ALPORTEL	-0,9645
SÃO JOÃO DA MADEIRA	-0,1270
SÃO PEDRO DO SUL	-0,5280
TAROUCÁ	-0,1983
TAVIRA	-0,7323
TERRAS DE BOURO	-0,0229
TOMAR	-0,3208
TONDELA	-0,4078
TORRES NOVAS	-0,3802
TORRES VEDRAS	-0,4619
TROFA	-0,0527
TÁBUA	-0,6979
VAGOS	-0,6398
VALE DE CAMBRA	-0,2697
VALENÇA	-0,5476
VALONGO	0,0602
VALPAÇOS	-0,6272
VENDAS NOVAS	-0,5503
VIANA DO CASTELO	-0,3676

**Tabela 3.32.** Estimativa dos coeficientes da variável concelho2 (13)

concelho2	
Variável	$\beta$
SÃO JOÃO DA PESQUEIRA	-0,9340
VILA NOVA DE FOZ CÔA	-0,7466
VILA VELHA DE RODÃO	-0,2187
VILA VERDE	-0,3638
VILA VIÇOSA	-0,3725
VIMIOSO	-1,0890
VINHAIS	-0,7521
WISEU	-0,3075
VIZELA	0,0204
VOUZELA	-0,2397
VILA NOVA DE CERVEIRA	-0,9106
REDONDO	-0,7656
SERPA	-0,5229
MOIMENTA DA BEIRA	-0,4901
ALFÂNDEGA DA FÉ	-1,3661
PENEDONO	-0,9340
TABUAÇO	-0,9340
MEDA	-0,7466
TRANCOSO	-0,7466
FIGUEIRA DE CASTELO RODRIGO	-0,7466

**Tabela 3.33.** Estimativa dos coeficientes da variável concelho2 (14)

concelho2	
Variável	$\beta$
PAMPILHOSA DA SERRA	-0,6657
REGUENGOS DE MONSARÁS	-0,7656
MAÇÃO	-0,7973
ALCOUTIM	-1,0102
CONSTÂNCIA	-0,7973
MOURÃO	-0,7656
CASTRO MARIM	-1,0102
CUBA	-0,3310
VIANA DO ALENTEJO	-0,3310
VILA FLÔR	-1,3661
GAVIÃO	-0,7973
MÉRTOLA	-0,5229
TORRE DE MONCORVO	-1,3661
ÁGUEDA	-0,3172
ÉVORA	-0,3872
ÍLHAVO	-0,4249
ÓBIDOS	-0,6504
BARRANCOS	-0,7656

**Tabela 3.34.** Estimativas dos coeficientes da variável *CVE\_TIPO\_GARAGEM*

CVE_TIPO_GARAGEM	
Variável	$\beta$
1	0,0000
2	-0,5289
3	-8,0022
4	-0,1445
Unknown	-7,9997

**Tabela 3.35.** Estimativa dos coeficientes da variável *idadeveic*

idadeveic	
Variável	$\beta$
<1ano	-0,2056
1 ano	-0,2393
2 a 3 anos	-0,1795
4 a 5 anos	-0,0937
6 a 7 anos	-0,0811
8 a 10 anos	-0,0653
+10 anos	0,0000

**Tabela 3.36.** Estimativa dos coeficientes da variável *NUM\_CC\_categ*

NUM_CC_categ	
Variável	$\beta$
0-500	-1,1870
500-1000	-0,0409
1000-1250	0,0161
1250-1500	0,0000
1500-1750	-0,0174
1750-2000	0,0064
2000-2250	-0,0077
2250-2500	-0,0948
2500-2750	0,0085
2750-3000	0,0610
3000-3250	-0,1221
3250-3500	0,0106
3500-3750	-0,3176
3750-4000	0,0278
+4000	-0,1623

**Tabela 3.37.** Estimativa dos coeficientes da variável *NUM\_TARA\_categ*

NUM_TARA_categ	
Variável	$\beta$
0-900	-0,0631
900-1500	0,0000
1500-2000	0,0239
2000-2500	0,0902
+2500	0,2320

Assim chegamos ao modelo final:

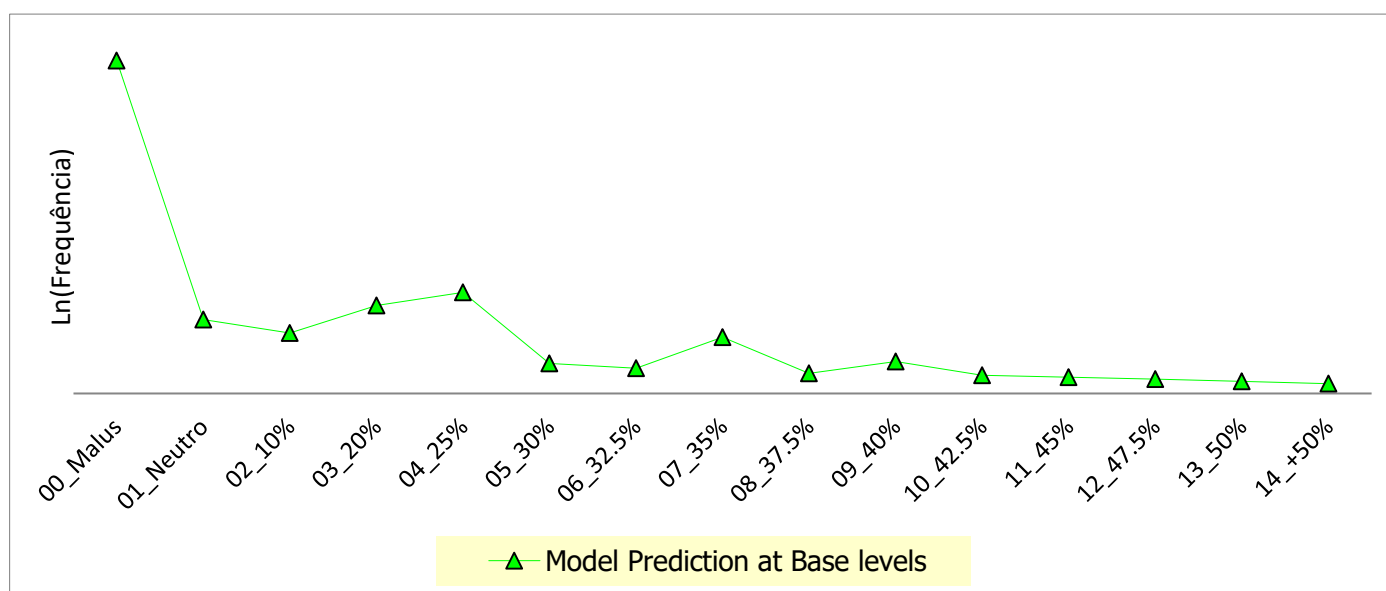
$$\ln(\text{frequência}) = \beta_0 + \sum_{i=1}^{15} \beta_i \times \text{bonus}_i + \sum_{i=1}^7 \beta_i \times \text{zonarede}_i + \sum_{i=1}^{47} \beta_i \times \text{marca}_i + \sum_{i=1}^4 \beta_i \times \text{codcomb}_i + \sum_{i=1}^4 \beta_i \times \text{pack}_i + \sum_{i=1}^2 \beta_i \times \text{cliente}_i + \sum_{i=1}^{278} \beta_i \times \text{concelho2}_i + \sum_{i=1}^4 \beta_i \times \text{fpagamento}_i + \sum_{i=1}^8 \beta_i \times \text{idadecond}_i + \sum_{i=1}^8 \beta_i \times \text{idadecart}_i + \sum_{i=1}^9 \beta_i \times \text{contrato}_i + \sum_{i=1}^7 \beta_i \times \text{idadeveic}_i + \sum_{i=1}^5 \beta_i \times \text{cve_tipo_garagem}_i + \sum_{i=1}^3 \beta_i \times \text{ano}_i + \sum_{i=1}^5 \beta_i \times \text{num_tara_categ}_i + \sum_{i=1}^{15} \beta_i \times \text{num_cc_categ}_i$$

**Equação 3.1.** Modelo Final

Para finalizar o capítulo foi necessário descobrir o público-alvo tendo em conta as variáveis explicativas do modelo final, ou seja, quais são as características do contrato e do cliente da seguradora em que a frequência de sinistralidade seja a menor possível, porque se existem menos sinistros menor será o prejuízo para a seguradora e assim maior será a rentabilidade da mesma.

Cada variável explicativa do modelo é analisada individualmente. Desta análise foi excluída a variável marca porque existem muitos segmentos desta variável, em que alguns têm pouca exposição e por isso não fazia sentido analisar cada segmento um por um.

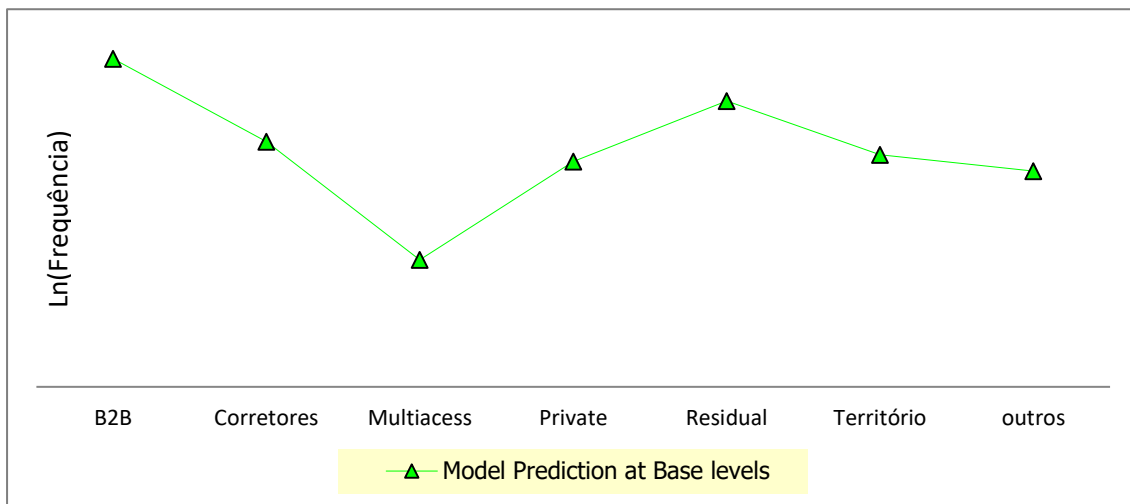
bonus:



*Gráfico 3.1. Aplicação do modelo na variável bonus*

Os segmentos da variável *bonus* com melhor frequência são os mais à direita que faz sentido porque estes são os clientes que devido à sua pouca sinistralidade lhes são dados bonus mais elevados. Pode-se afirmar que o cliente alvo tem de ter um bonús igual ou superior a 42.5%.

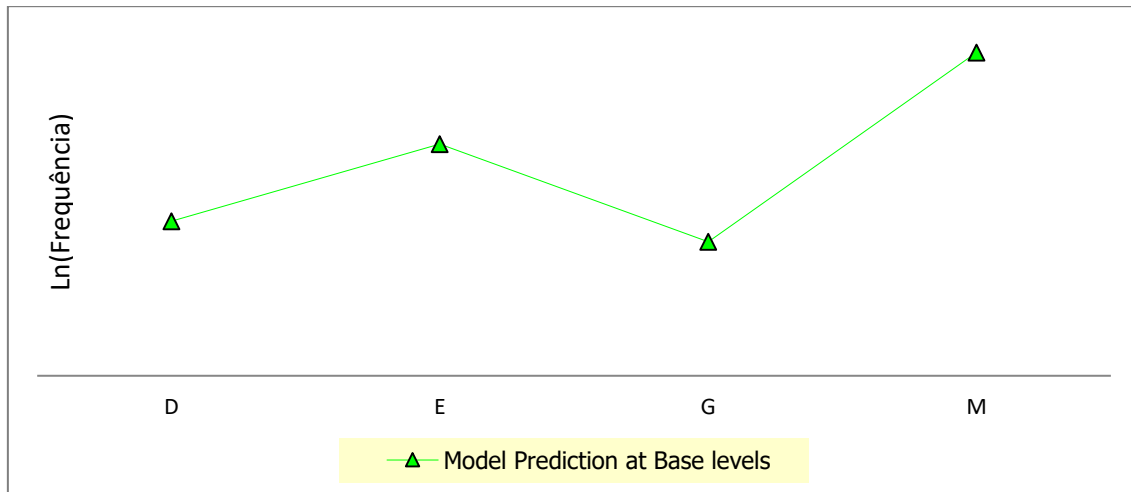
zonarede:



*Gráfico 3.2 Aplicação do modelo na variável zonarede*

O segmento da variável zonarede que o modelo determina que tem a menor frequência de sinistralidade é o multiacesso. O multiacesso são as lojas fixas que apenas vendem seguros da AXA/Ageas e que estão espalhadas pelo país.

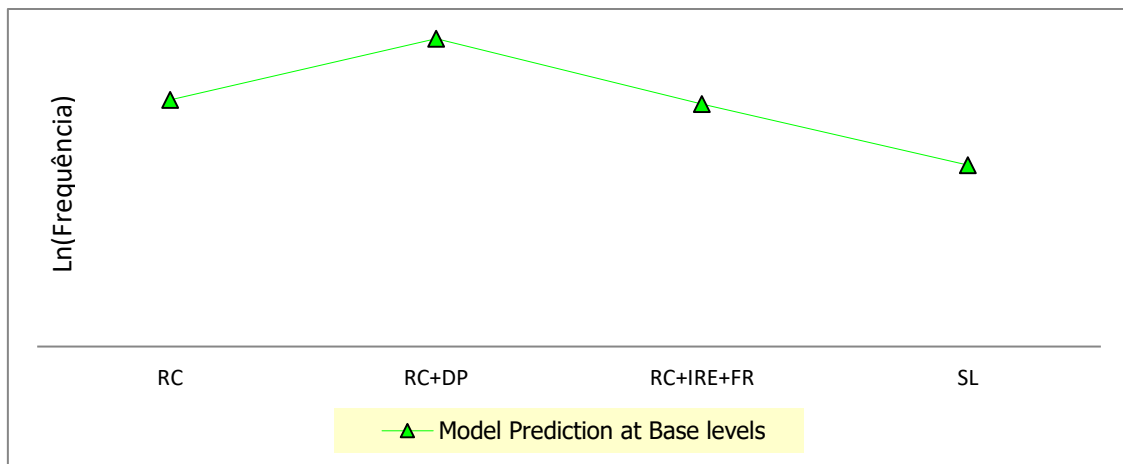
codcomb:



*Gráfico 3.3 Aplicação do modelo na variável codcomb*

Os segmentos da variável codcomb que o modelo determina que tem a menor frequência de sinistralidade são o D e o G.

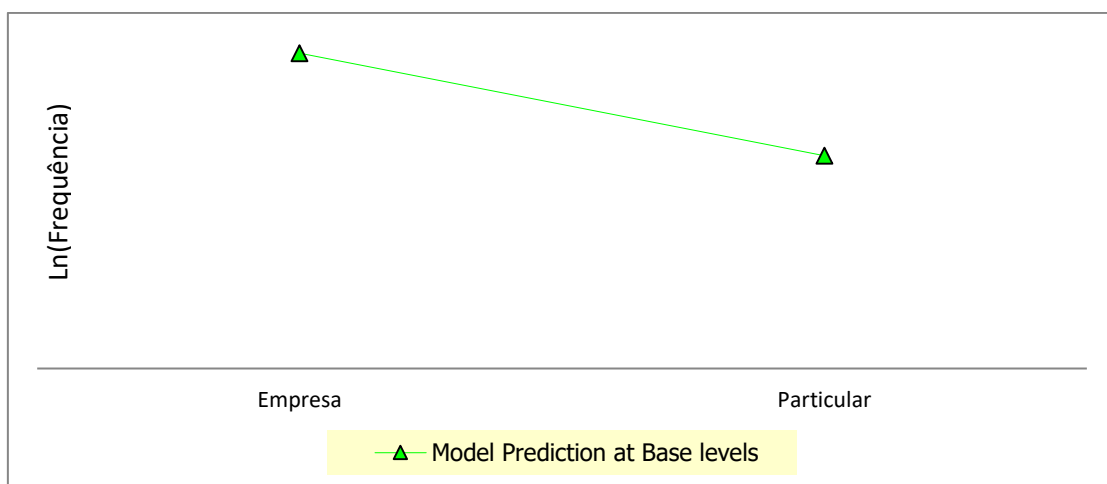
pack:



*Gráfico 3.4 Aplicação do modelo na variável pack*

Os segmentos da variável pack que o modelo determina que tem a menor frequência de sinistralidade são o RC+IRE+FR e o SL. O serviço à lista (SL) tem muito pouca exposição por ser uma cobertura muito especial e se escolhessemos este segmento para o público-alvo não seria lucrativo para a companhia porque teria poucos clientes. O segmento RC+IRE+FR corresponde ao pack que contém as coberturas de responsabilidade civil, incêndio, raio e explosão e ainda a cobertura de furto ou roubo. Como este pack tem uma frequência baixa relativamente aos restantes segmentos, é o escolhido para o cliente-alvo.

cliente:

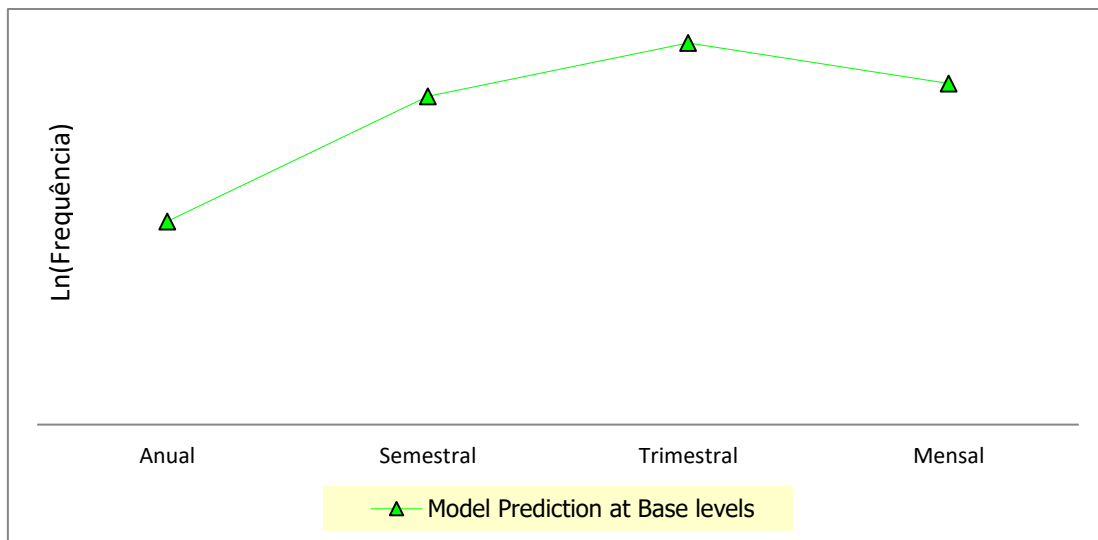


*Gráfico 3.5 Aplicação do modelo na variável cliente*



Um cliente particular tem uma menor frequência que um cliente empresa porque por norma um cliente empresa tem muitos mais carros em circulação e durante um período de tempo maior do que um cliente particular. O segmento escolhido para o cliente-alvo é o particular.

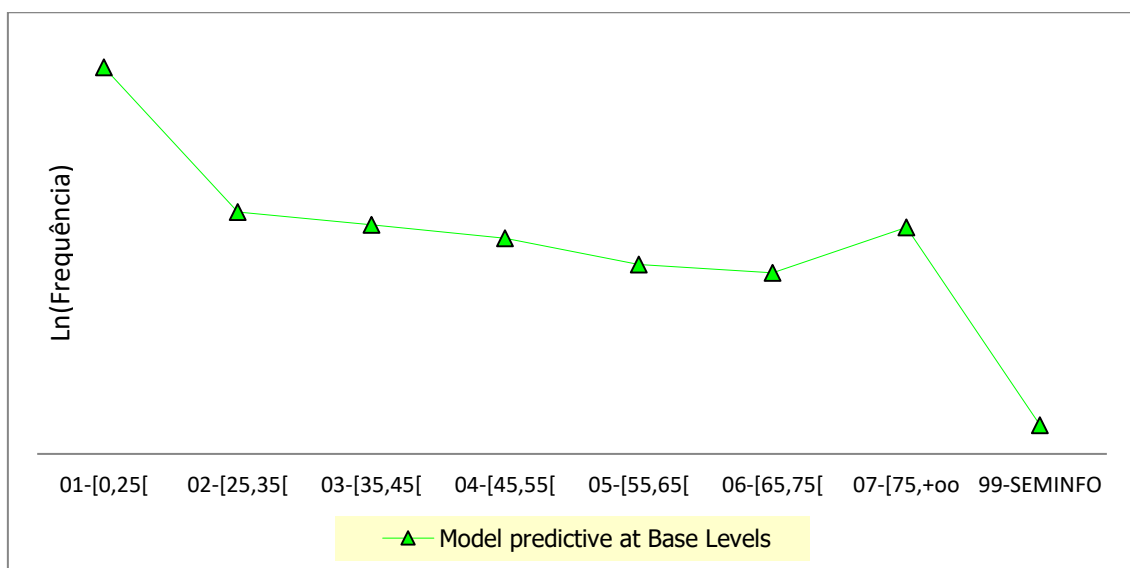
*fpagamento:*



**Gráfico 3.6** Aplicação do modelo na variável *fpagamento*

O segmento da variável *fpagamento* que o modelo determina que tem a menor frequência de sinistralidade é a forma de pagamento anual, em que o cliente paga o seu seguro por inteiro em apenas uma prestação.

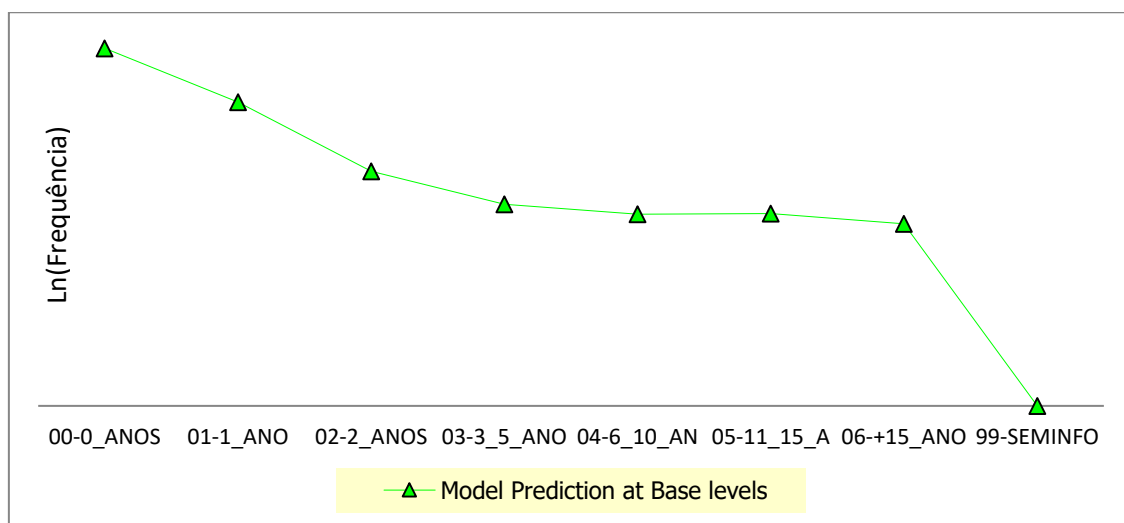
*idadecond:*



**Gráfico 3.7** Aplicação do modelo na variável *idadecond*

Os segmentos da variável idadecond que o modelo determina que tem a menor frequência de sinistralidade são entre o segundo e o quinto segmento. Esta variável correspondem à idade do condutor em que os segmentos com melhor frequência equivalem aos condutores com idades compreendidas entre 25 e 74 anos, o que pode ser explicado pelo fato que condutores com idades inferiores a 25 anos têm pouca experiência a conduzir o que contribui para um aumento da sinistralidade porque não se sentem tão à vontade, e condutores com idades superiores a 74 anos porque já têm uma certa idade e como tal as suas capacidades motoras e de reação já estão um pouco debilitadas o que contribuiu também para uma alta sinistralidade.

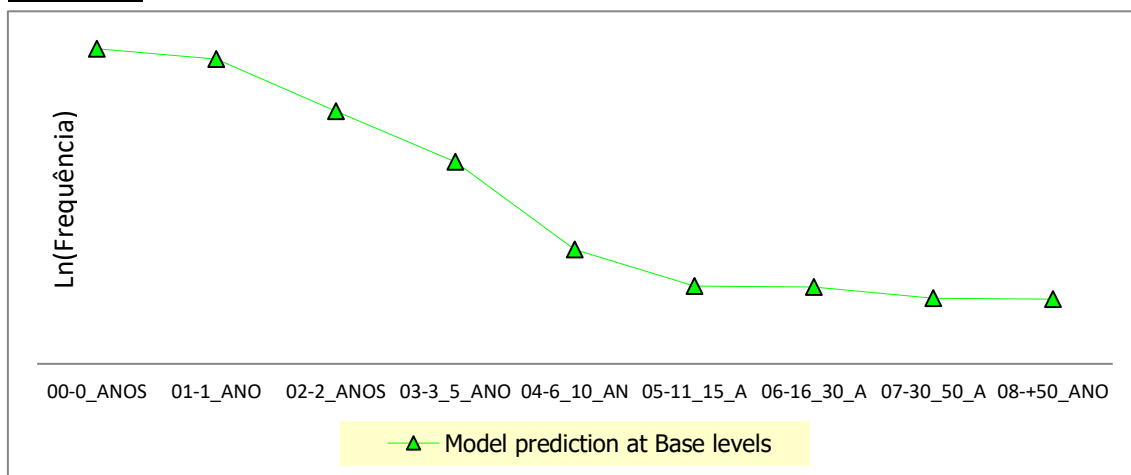
idadecart:



*Gráfico 3.8 Aplicação do modelo na variável idadecart*

Os segmentos da variável idadecart que o modelo determina que têm a menor frequência de sinistralidade são a partir do quarto segmento inclusivo. Esta variável corresponde à idade da carta do condutor em que os segmentos com melhor frequência equivalem aos condutores com experiência de condução superior a 3 anos. Pelo fato dito acima de que clientes muitos novos têm uma frequência alta por terem pouca experiência de condução, o mesmo acontece aqui porque os primeiros 3 segmentos correspondem a condutores com pouquíssima experiência e como tal não têm a sensibilidade e o pouco tempo de reação que condutores mais experientes têm.

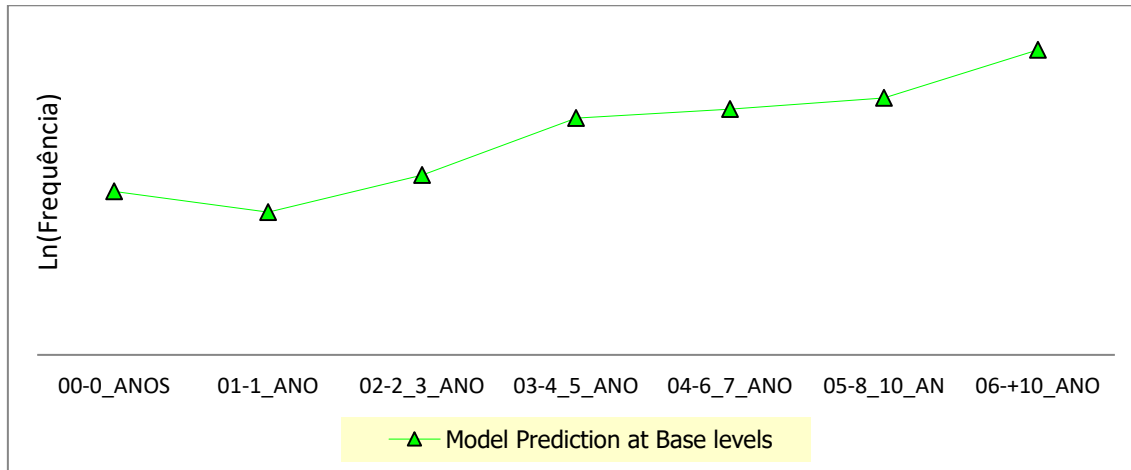
contrato:



*Gráfico 3.9 Aplicação do modelo na variável contrato*

Os segmentos da variável *contrato* que o modelo determina que têm a menor frequência de sinistralidade são a partir do sexto segmento. Esta variável corresponde à quantidade de anos em que a apólice está em vigor. Por isso pelo modelo pode-se afirmar que os melhores clientes são os que estão na companhia há mais de 11 anos.

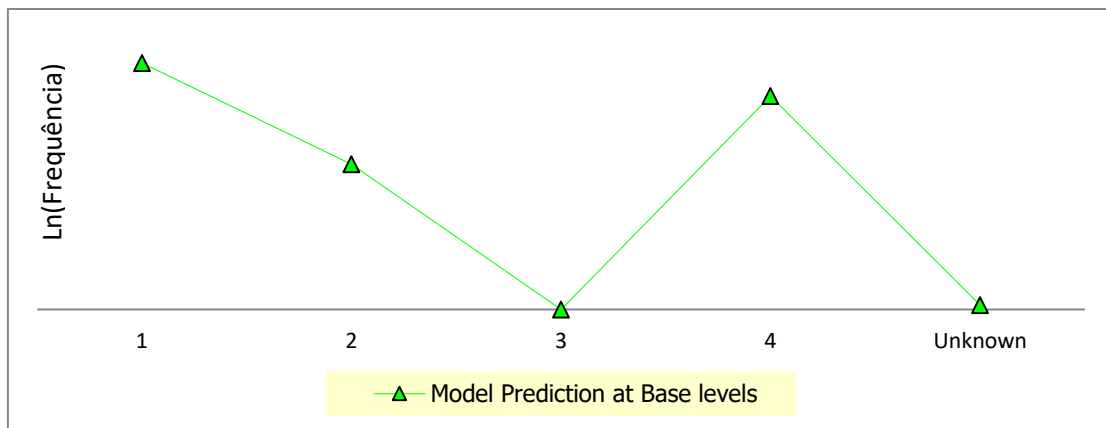
idadeveic:



*Gráfico 3.10 Aplicação do modelo na variável idadeveic*

Os segmentos da variável *idadeveic* que o modelo determina que têm a menor frequência de sinistralidade são os dois primeiros segmentos. Esta variável corresponde à idade do veículo seguro na apólice, e como tal veículos novos têm mais cuidados por parte do condutor e assim são menos susceptíveis a acidentes. Por isso pelo modelo pode-se afirmar que os melhores clientes são os que têm carros com menos de 1 ano inclusivé.

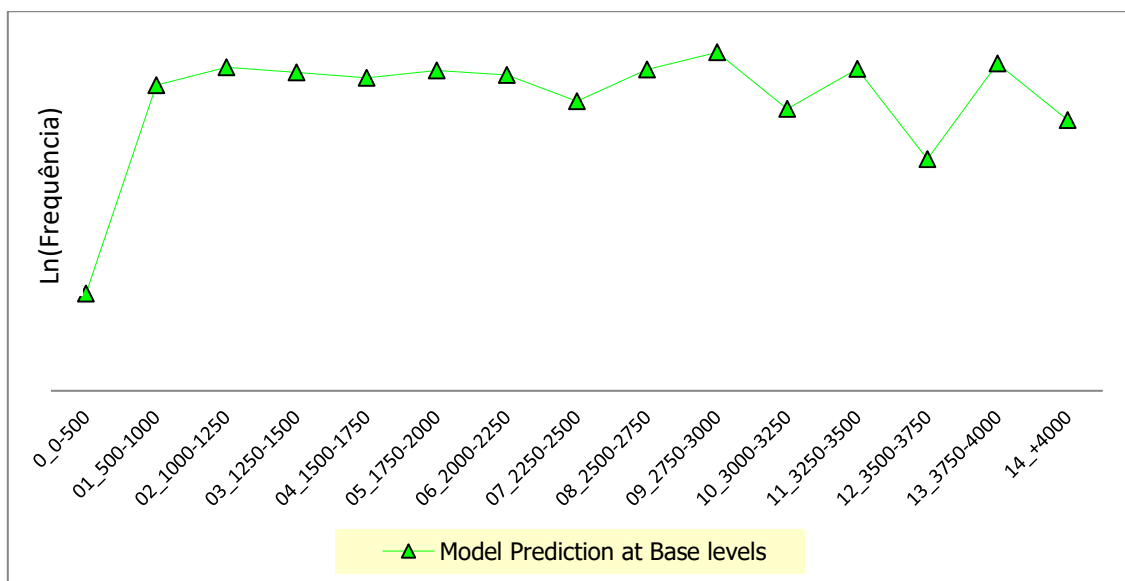
### CVE TIPO GARAGEM:



**Gráfico 3.11** Aplicação do modelo na variável *CVE\_TIPO\_GARAGEM*

O segmento da variável *cve\_tipo\_garagem* que o modelo determina que tem a menor frequência de sinistralidade são os clientes que só têm garagem durante o dia.

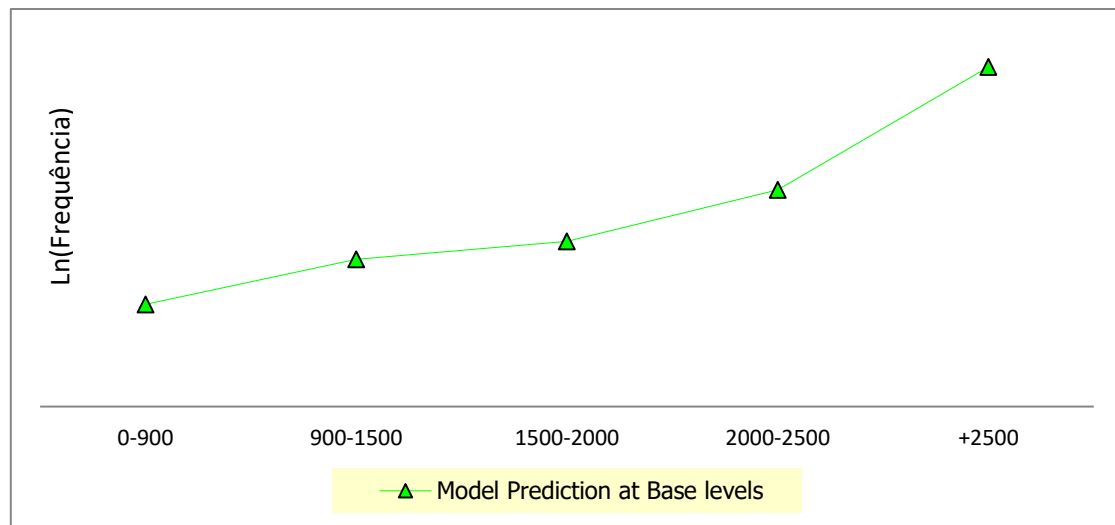
### NUM CC CATEG:



**Gráfico 3.12** Aplicação do modelo na variável *NUM\_CC\_CATEG*

Para a variável *num\_cc\_categ* o modelo determina que os veículos com pouca cilindrada são os que têm a menor frequência de sinistralidade.

### NUM\_TARA\_CATEG:



*Gráfico 3.13 Aplicação do modelo na variável NUM\_TARA\_CATEG*

Para a variável *num\_tara\_categ* o modelo determina que os veículos com pouca tara são os que têm a menor frequência de sinistralidade

Assim o cliente alvo da companhia para o seguro automóvel é um cliente com:

- bonús acima de 42.5%;
- rede de agentes multiacesso;
- tipo de combustível D ou G;
- pack com as coberturas de RC, IRE e FR;
- particular;
- forma de pagamento do prémio anual;
- entre 25 e 74 anos;
- com mais de 3 anos de carta;
- que já sejam clientes há mais de 11 anos;
- com veículos novos (1 ano ou menos);
- com garagem apenas durante o dia;
- com veículos com cilindrada abaixo dos 500;
- e com veículos com tara abaixo de 2000;

## Conclusão/Discussão

A realização do estágio curricular traduziu-se numa experiência profissional muito enriquecedora, pois permitiu-me contactar com o mundo profissional, ou seja, consegui implementar os conhecimentos adquiridos ao longo do mestrado na realidade empresarial mais concretamente no mercado segurador.

A escolha de realizar um relatório de estágio como projeto final foi no primeiro momento a escolha ideal sendo a sua principal razão a oportunidade de ingressar no mercado de trabalho, mais especificamente na minha área. Nos tempos em que nos encontramos sabe-se que é cada vez mais difícil encontrar emprego e estabilidade em algo que nos especializamos, e foi por isso que na altura escolhi a seguradora AXA Portugal (atual Ageas Seguros) porque é uma empresa multinacional de referência.





Sendo o mercado segurador um negócio muito competitivo as seguradoras procuram encontrar formas de se distinguirem da sua concorrência. Uma das principais formas de o fazerem é através da tarificação dos seus produtos, e foi neste ponto que consistiu o projeto, em que o principal objetivo foi não só criar zonas tarifárias como também perceber como é que as características do contrato/cliente se comportam, ou seja, quais são as variáveis que explicam a frequência de sinistralidade e quais são os segmentos dentro dessas variáveis que são o nosso público-alvo.

Através deste projeto pode-se concluir que a zona tarifária com baixo risco é representado pelo interior norte e centro, com médio-baixo risco a zona do Alentejo e Algarve, com médio e médio-alto risco os concelhos circundantes às grandes cidades populacionais e por fim com alto risco os concelhos de Lisboa, Porto e Vila Nova de Gaia.

As variáveis de interesse do modelo final obtido são:

- |              |                    |
|--------------|--------------------|
| ✓ bonus      | ✓ idadecond        |
| ✓ zonarede   | ✓ idadecart        |
| ✓ marca      | ✓ contrato         |
| ✓ codcomb    | ✓ idadeveic        |
| ✓ pack       | ✓ CVE_TIPO_GARAGEM |
| ✓ cliente    | ✓ NUM_CC_categ     |
| ✓ concelho2  | ✓ NUM_TARA_categ   |
| ✓ fpagamento | ✓ ano              |

Sendo que o cliente-alvo é o cliente com as seguintes características:

-  bonús acima de 42.5%
-  rede de agentes multiacesso
-  tipo de combustível D ou G
-  pack com as coberturas de RC, IRE e FR

- 👉 particular
- 👉 forma de pagamento do prémio anual
- 👉 entre 25 e 74 anos
- 👉 com mais de 3 anos de carta
- 👉 que já sejam clientes há mais de 11 anos
- 👉 com veículos novos (1 ano ou menos)
- 👉 com garagem apenas durante o dia
- 👉 com veículos com cilindrada abaixo dos 500
- 👉 e com veículos com tara abaixo de 2000

Relativamente às potenciais limitações com que me deparei, é de destacar o uso do Excel pois a base de dados interna que utilizei, que consiste numa ferramenta de trabalho indispensável, era muito grande para ser analisada em Excel o que tornou o meu trabalho de análise muito pesado e demorado.

Por fim é de salientar a excelente oportunidade que a empresa me proporcionou, e que foi uma satisfação enorme realizar este projeto.

## Referências Bibliográficas

Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., & Thandi, N. (2007). *A Practitioner's Guide to Generalized Linear Models: A foundation for theory, interpretation and application*, 3ª Edição.

Amaral Turkman, M.A. e Silva, G (2000). Modelos Lineares Generalizados – da Teoria à Prática, Editions SPE, Lisboa

Apontamentos da professora Teresa Alpuim para apoio à disciplina Modelos Lineares do curso Matemática Aplicada à Economia e Gestão.

Documentos de apoio ao curso “Estatística e atuariado para não atuários” fornecidos pela Associação Portuguesa de Seguradores.

M Yuan, Y Lin (2006) Journal of the Royal Statistical Society: Series B (Statistical Methodology) - Model selection and estimation in regression with grouped variables, Royal Statistical Society.

Garraio, J. B. (Setembro de 2015). Modelação da Taxa de Anulação no Seguro Automóvel. FCUL-UL.

Pousinho, A. P. (Dezembro de 2013). Modelo Linear Generalizado Tweedie aplicado ao cálculo de provisões para sinistros. ISEG-UTL

## Referências Sitográficas

História dos seguros em Portugal. Disponível: <http://historiadoseguro.com/historia-de-seguros-portugal/> [Acesso em: 2016/04]

AXA Portugal. Disponível: <https://www.axa.pt/> [Acesso em: 2016/04]

Grupo AXA. Disponível: <http://www.axa.com/en/group/> [Acesso em: 2016/04]

Plataforma digital “E-learning” disponível para os colaboradores da empresa AXA-Portugal/ Ageas Seguros [Acesso em: 2016/04]